SENSORY INFORMATION PROCESSING AND SYMBOLIC COMPUTATION

Thomas G. Stockham, Jr.

Utah University

Prepared for:

Defense Advanced Research Projects Agency

June 1975

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER / 2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) SENSORY INFORMATION PROCESSING AND SYMBOLIC COMPUTATION	5. TYPE OF REPORT & PERIOD COVERED Semi-Annual 1 Jan 75 to 30 Jun 75
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s) DAHC15-73-C-0363
9. PERFORMING ORGANIZATION NAME AND ADDRESS Computer Science Department University of Utah Salt Lake City, Utah 84112	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS ARPA Order Number: 2477
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, Virginia 22209	12. REPORT DATE June 1975
	13. NUMBER OF PAGES 138
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

image deblurring, LPC spectral matching algorithms, audio, word recognition, speech intelligibility, REDUCE, LISP, sparse matrices, factored polynomial algebra, shaded pictures, three-dimensional objects

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

SENSORY INFORMATION PROCESSING--This research uses digital computation to investigate processes, both linear and nonlinear, for the filtering, restoration enhancement, bandwidth reduction, distortion immunization and analysis of both visual and auditory information.

Section 1-3 reports additional results in image deblurring that extend results reported in the last semi-annual report (1 July 74 - 31 Dec 74). Technical reports covering this work are planned. Our initial efforts to extend these results into color are reported in Section 4. (Cont'd)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Several items of substantial progress are reported in the speech area, notably Sections 5, 6, 8, and 10. Section 5 reports the results obtained from basic LPC spectral matching algorithms that have been developed and implemented as tools for all researchers at this installation. Section 6 reports continuing results in improving quality of processed audio. Section 8 reports final results of isolated word recognition work, and Section 9 reports some extentions contemplated along similar lines. Section 10 reports ongoing results in methods for improving speech intelligibility.

Sections 7 and 11 report results that have been in development some time. These items have not previously been reported, and have reached the state of having been documented during this period. Separate ARPA technical reports on these items are not now contemplated.

Section 12 outlines an area of future interest concerning the estimation of the phase of a speech wave.

SYMBOLIC COMPUTATION--The goal of this research is the production of efficient programs for algebraic and symbolic computation related to scientific research. The long range objective is the development of a completely automatic algebraic programming system which can be moved easily from one computer to another.

Section 1 reports continuing progress in mode analyzing for REDUCE. Section 2 reports ongoing results to improve the IBM 360 LISP which is basic to the REDUCE transportability. Section 3 reports continuing efforts to improve implementation of space matrices and factored polynomial algebra. Specific new packages that have been implemented are described in Section 4.

GRAPHICS--This research effort plans to complete the construction and demonstration of a system which makes shaded pictures of computer models of three-dimensional illuminated objects, and to conduct research to find substantially improved modeling techniques for dynamically changing object collections.

Several items of research reached fulfillment during the period as reported in Sections 1, 2, and 4. Related technical reports are in process. Section 3 indicates ongoing work.

2

SENSORY INFORMATION PROCESSING AND SYMBOLIC COMPUTATION

1 January 1975 THROUGH 30 June 1975

Semi-Annual Technical Report

Contractor: University of Utah
Contract Number: DAHC15-73-C-0363
Effective Date: 1 July 1973
Expiration Date: 30 September 1975
Amount of Contract: $2,275,000.00
Project Code: 3D30

Principal Investigator: Dr. Thomas G. Stockham, Jr.
Telephone: (801)581-8224

Contracting Officer: Mr. Edgar S. Allen
DSSW

Approved for public release;
distribution unlimited.

Sponsored by
Defense Advanced Research Projects Agency
ARPA Order Number 2477

UTEC-CSc-76-008
Computer Science

TABLE OF CONTENTS

REPORT SUMMARY

Sensory Information Processing

This research uses digital computation to investigate processes, both linear and nonlinear, for the filtering, restoration, enhancement, bandwidth reduction, distortion immunization and analysis of both visual and auditory information.

Section 1 - 3 reports additional results in image deblurring that extend results reported in the last semi-annual report (1 July 74 - 31 Dec 74). Technical reports covering this work are planned. Our initial efforts to extend these results into color are reported in Section II-4.

Several items of substantial progress are reported in the speech area, notably Sections 5, 6, 8, and 10. Section 5 reports the results obtained from basic LPC spectral matching algorithms that have been developed and implemented as tools for all researchers at this installation. Section 6 reports continuing results in improving quality of processed audio. Section 8 reports final results of isolated word recognition work, and Section 9 reports some extentions contemplated along similar lines. Section 10 reports ongoing results in methods for improving speech intelligibility.

Sections 7 and 11 report results that have been in development some time. These items have not previously been reported, and have reached the state of having been documented during this period.

Separate ARPA technical reports on these items are not now contemplated.

Section 12 outlines an area of future interest concerning the estimation of the phase of a speech wave.

Symbolic Computation

The goal of this research is the production of efficient programs for algebraic and symbolic computation related to scientific research. The long range objective is the development of a completely automatic algebraic programming system which can be moved easily from one computer to another.

Section 1 reports continuing progress in mode analyzing for REDUCE. Section 2 reports ongoing results to improve the IBM 360 LISP which is basic to the REDUCE transportability. Section 3 reports continuing efforts to improve implementation of space matrices and factored polynomial algebra. Specific new packages that have been implemented are described in Section 4.

Graphics

This research effort plans to complete the construction and demonstration of a system which makes shaded pictures of computer models of three-dimensional illuminated objects, and to conduct research to find substantially improved modeling techniques for dynamically changing object collections.

Several items of research reached fulfillment during the period as reported in Sections 1, 2, and 4. Related technical reports are in process. Section 3 indicates ongoing work.

II. RESEARCH ACTIIVIITIIES

SENSORY INFORMATION PROCESSING

Section 1

Restoration of Blurred Images
Brent Baxter

This section discusses the continuation of work reported under the same title in the previous semi-annual report (1 July 1974 to 31 December 1974). That report noted in some detail a model for predicting brightness perception which was able to adapt to the spacial frequency content of an image. A little reflection suggested a similarity between blurred photographs and images containing weak high frequency texture, and a system capable of accentuating weak texture was devised which proved euperior to previous methods for restoring blurred photographs.

1.1 Restoration of Blurred Images

The ideae upon which this section is based first came to mind by observing that each AGC element in the frequency selective model has associated with it a criterion for determining when the signal in its channel is too weak and must be amplified. This means that the local power spectrum * is being adjusted to some previously specified prototypical value. These ideas led to the deblurring mechanism in Figure 1.1.

---

* The idea of a local power spectrum is similar to the short time spectrum on which speech spectrograms are based.

The method will be described as concisely as possible despite the fact that in certain respects its mathematical foundations are rather subtle. Following this discussion, some comparisons between the deblurring method and the frequency selective model will be made which suggest a limited deblurring capability in human vision.

Blurring caused by camera motion or an out of focus lens are processes that combine a clear image with a blur impulse response by convolution. An out of focus blur has a cylindrical impulse response and a motion blur has a fence-like impulse response. Stylized examples are shown in Figures 1.3 and 1.17. The process of deconvolution [1] * is the process of removing one of the components (the blur) by first mapping the blurred image into a space in which the two components are added rather than convolved, and then the blur information is removed by linear filtering **.

Several problems stand in the way of what might seem otherwise to be a straightforward task:

---

* See references for Sections 1, 2 and 3 after Section 3.

** There is a terminology associated with this style of signal processing in which the filtering just described is known as liftering, and it is implemented by multiplication in the so-called cepstral domain. The independent variable in this domain is called quefrency and is described as being short or long rather than high and low as is the case for the frequency domain. Using these terms Figure 1.1 is really a long pass lifter. For the purposes of this report the more familiar terms Fourier transform, logarithm and filter will be used to describe these processes.

(1) The first complication involves noise-like fluctuations, known as film grain, introduced when images are recorded on photographic film. If the restoration is to be acceptable these random fluctuations must not be amplified excessively by the restoration process.

(2) The second difficulty arises because space-limited blurs of the type illustrated above completely eliminate image energy at certain spatial frequencies making an exact restoration impossible even in the absence of film grain noise. However, if there is no noise and only a few frequencies are eliminated, their absence goes almost unnoticed. Figures 1.2-1.6 illustrate this.

(3) The third problem is that approximate aperiodic convolutional inverses for space-limited blurs are often nonzero * over large domains **.

(4) The fourth problem is caused because the blurred image is truncated abruptly at the edges of the film by the camera's film holder. Special edge treatments are required to suppress artifacts generated by these edges.

---

* See the footnote in Appendix B for an interpretation of the term nonzero.

** See Appendix A for an illustration of this difficulty.

Item number three can be particularly troublesome because restoration filters for use with aperiodic convolution often have very long impulse responses, and arbitrarily truncating them for compuational convenience causes severe ghost-like echoing in the restored image. Figures 1.7 and 1.8 illustrate this problem for the artificial blur. Traditional attempts to remedy the situation by windowing have proved to be only partially successful. See Cannon [2] page 34 and Cole [3] page 53 for examples of echo-like artifacts introduced in blur removal schemes by impulse response truncation.

An alternative to windowing is to take advantage of the fact that both the blurred input image and restored image are usually of finite size (often the same size) and a convolutional inverse can be truncated without introducing echoes provided that its domain is adequate ☆.

The approach to the truncation problem taken here relies on the limited spatial extent of most blur impulse responses which allows substitution of periodic ☆☆ for aperiodic convolution. The blur systems considered here completely eliminate signal energy at certain spatial frequencies and severly attenuate it at others nearby. In attempting to amplify these frequencies adequately, the system is constrained by the highpass filter so that noise is not

---

☆ See Appendix B for a discussion of constraints on the impulse response length imposed by this technique. This approach may be impratical because of the large computational effort involved in obtaining the convolutional inverse prior to truncation.

☆☆ Appendix C shows how the effects of circular convolution may be obtained by filtering the log spectrum.

amplified excessively. The next four figures should help illustrate this issue. Success of this substitution is due to the fact that both kinds of convolution give identical results, except possibly near image boundaries. Note that results identical with Figure 1.4 are obtained if Figure 1.2 is extended with its boundary value, combined with Figure 1.3 by aperiodic convolution and the result truncated to the original size.

Figure 1.9 shows the frequency response associated with the artificial blur referred to earlier. Frequencies where this function is zero are the same ones where its logarithm (Figure 1.10) increases negatively without bound. It is the task of the highpass filter, Figure 1.11, to remove as much of this signal as possible while preserving regions of large negative value to prevent excessive amplification of noise. In Figure 1.12, the blur log spectrum * has been almost entirely removed except in the region of potentially noise dominated frequencies. This has the effect of restoring the image while preventing excessive amplification of noise.

This same filtering effect is obtained if image information is present along with the blur, but part of the image is removed by the filter and must be reinserted in the form of an equalization signal like the one in Figure 1.13. An equilization signal which is properly prepared can be used in restoring a variety of blurred images. One way to prepare an equalization signal is to lowpass

---

* Log spectrum is used to mean the real part of the complex logarithm of the discrete Fourier transform.

filter the log spectrum of a similar, unblurred image using a filter whose frequency response is related to that of the highpass filter as follows:

$$LPF(\omega) = 1(\omega) - HPF(\omega)$$

Filters of this type are called complementary filters.

Phase information of the blur system is not present in Figure 1.10 or Figure 1.12 and no method of removing it has been described so far. In the case of the artificial blur, the phase is known exactly and subtracting it is straightforward. For real blur systems encountered in the field, a method such as the one by Cannon [2] may be used to estimate the phase of the blur after which it may be subtracted. Figure 1.14 shows the phase associated with the artificial blur.

In Figure 1.15, the artificially blurred image is restored using the techniques described above. There are two reasons why the retoration is not identical with the one in Figure 1.6. First, the effects of noise are anticipated but none is present. This capacity for dealing with noise is not needed but it constrains the restoration as if it were needed. Second, the exact details of the blur log spectrum were unknown to the restoration system. This is known as "blind deconvolution."

Applying these ideas to images blurred in the field requires a special edge treatment to make the image have the same value along its boundary as shown in Figure 1.19. The blurred image is treated

as though it has been replicated periodically and a special interpolation process * is invoked near image boundaries to simulate periodic blurring across the boundaries. In this setting, questions regarding convolutional inverses having large spatial extent do not arise since the inverse is periodic and must be determined only for a single period. If the edge treatment is successful, no further attention is required with regard to problem four (truncation after blurring).

Figures 1.16-1.22 give an idea of the degree to which a successful restoration of images blurred in the field is possible using this method. Note how the small text in Figure 1.19, which was blurred beyond legibility, has been made legible in the restored image of Figure 1.21. Also, the outline of the sign is much more clearly defined in both Figures 1.18 and 1.21. The speckled appearance of Figures 1.18 and 1.21 is due to amplification of film grain noise and may be traded for sharpness by adjusting the highpass filter cutoff frequency. A minor artifact, barely perceptible in Figure 1.21, is a shadow manifesting itself as a dark copy of the letter "c" to the left of the word convention. This may

---

* The interpolation process consists of the following:

a. Removing the average value
b. Multiplying the image by unity everywhere except near edges and there by half of a Hanning window. Figure 1.23 illustrates this windowing operation.
c. Reinserting the average value.

This scheme simulates blurring across image boundaries.

be due to regularly spaced harmonics that are missing from the restored image or perhaps errors in estimating the phase of the blur.

Several similarities between the frequency selective model of vision (Figure 1.24) and the deblurring mechanism (Figure 1.1) suggest a limited deblurring capability in human vision. The Fourier transform is often naively understood in terms of a bank of bandpass filters similar to those in Figure 1.24, a principal difference being the lower frequency resolution of the filters in the visual model. The AGC elements in Figure 1.24 were implemented using Stockham's method [10] for separating multiplied signals and is similar to the logarithm, highpass filter, and exponential in the deblurring mechanism. One may view the equalization signal of Figure 1.1 as providing a fixed gain for each frequency channel. There are only two places where the analogy breaks down. There is no logarithmic stage operating on the blurred image prior to the Fourier transform in Figure 1.1, and the visual model has no provision for correcting phase reversals introduced by the blur. For images of low dynamic range, such as the sign in Figure 1.16, the logarithm is approximately linear and its absence is not critical. Failure to correct phase reversals is a significant difference between the two systems and is a major limitation on the ability of the visual system to restore blurred images.

The existence of this deblurring capability in human vision might help explain why it is difficult to make improvements in blurred photographs that are as striking as might be expected from

physical considerations.

1.2 Summary

In the previous semiannual report, experimental evidence favoring a frequency selective model of brightness perception has been described with emphasis on the model's texture equalization property. This property follows from electrophysiological as well as psychophysical experiments and forms the basis of an enhancement process where changes in contrast are made on the basis of texture strength as well as texture size. The successful method for removing unknown photographic blurs based on this property described here is very similar in its organization to that of the visual model. As a result of this similarity, a limited capacity for restoring blurred images is claimed for the visual system.

In the following section, a class of edge related illusions is considered in relation to a rather different model of brightness perception.

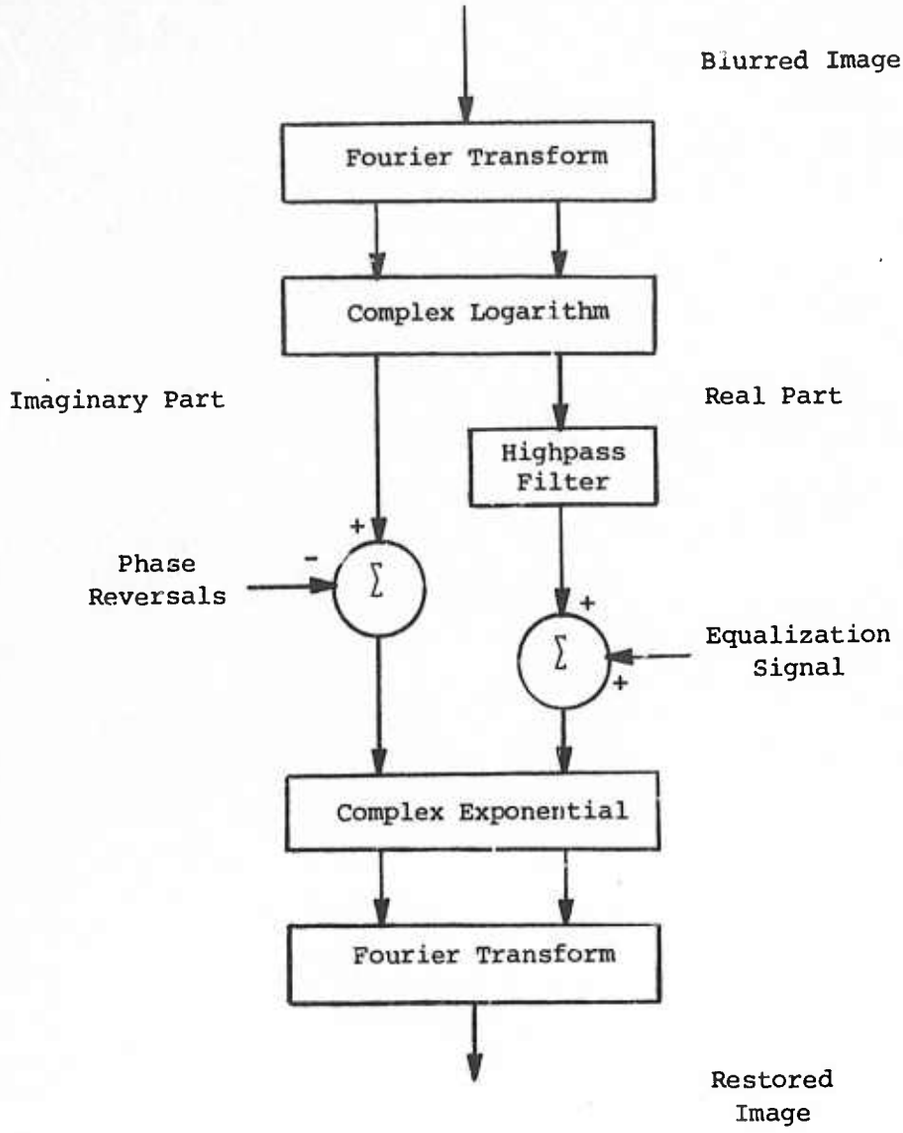


Figure 1.1 -- A deblurring mechanism suggested by the texture equalization property.

Figure 1.2 -- Artificially created test image.

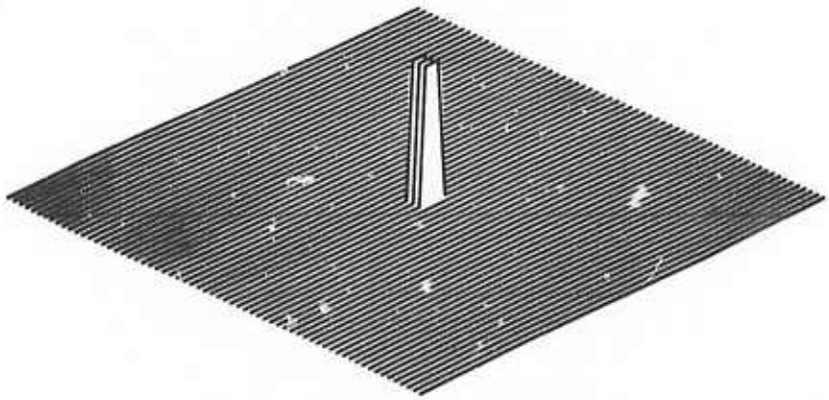


Figure 1.3 -- Artificial blur impulse response.

Figure 1.4 -- Test image artificially blurred using circular convolution (noise free).

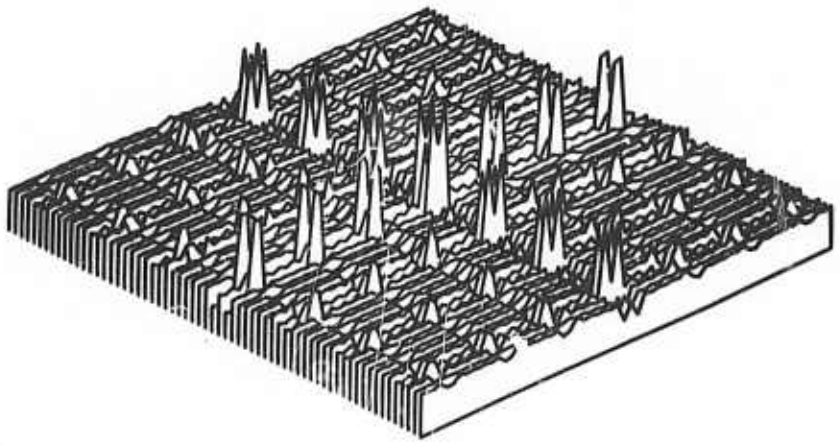


Figure 1.5 -- Convolutional inverse modified to avoid infinite gain at spatial frequencies where the blur system has zero gain.

Figure 1.6 -- Artificial image restored by inverse filtering.

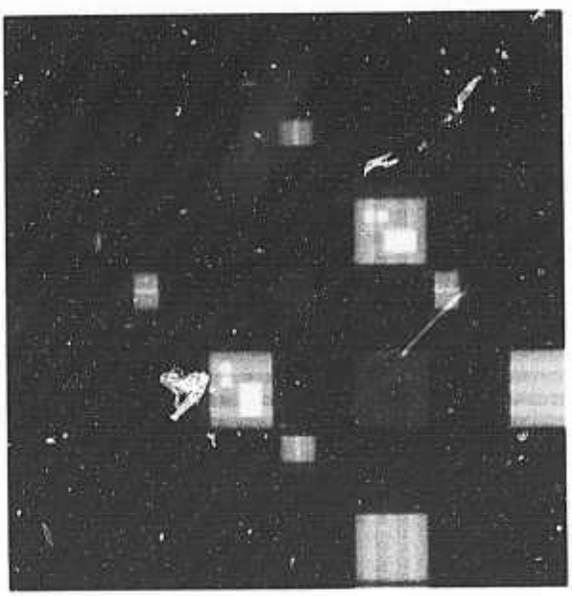


Figure 1.7 -- Artificial image restored using a convolutional inverse truncated to half its original size. Truncation caused large errors in the average value requiring manual corrections to make the image printable.

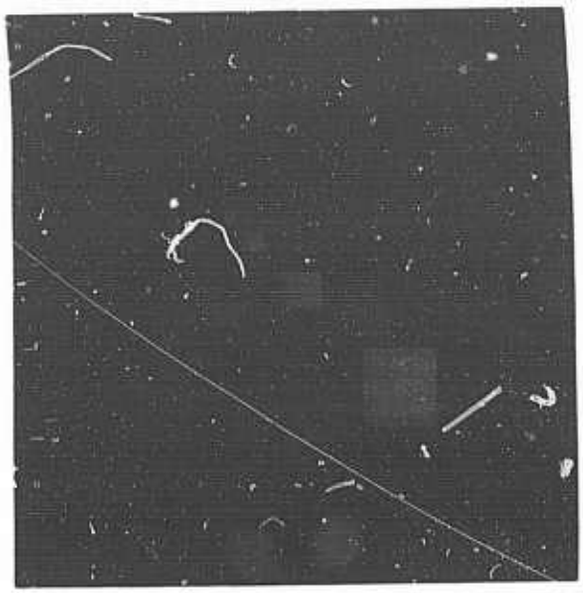


Figure 1.8 -- Same as Figure 1.7 except the truncated inverse impulse response was windowed to reduce echoing.

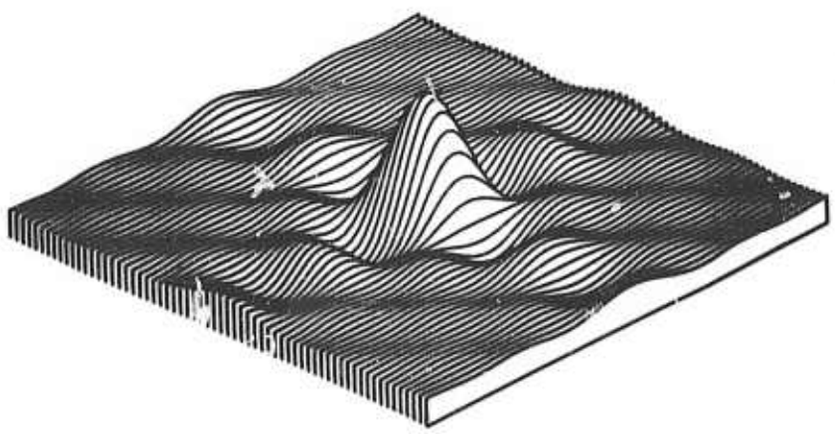


Figure 1.9 -- Blur frequency response.

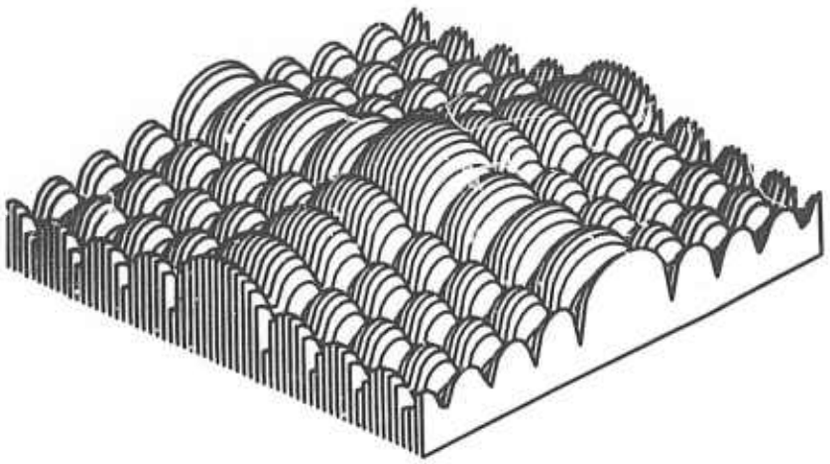


Figure 1.10 -- Log spectrum of the blur.

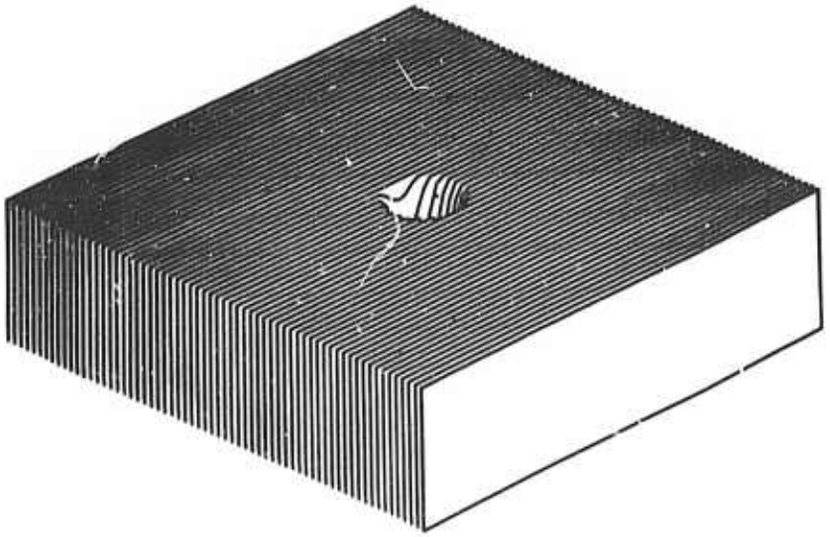


Figure 1.11 -- Highpass filter frequency response.

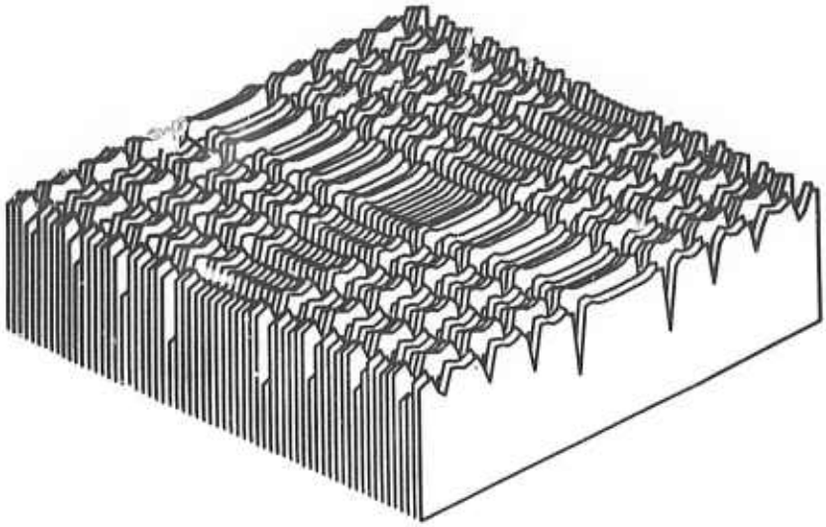


Figure 1.12 -- Filtered log spectrum.

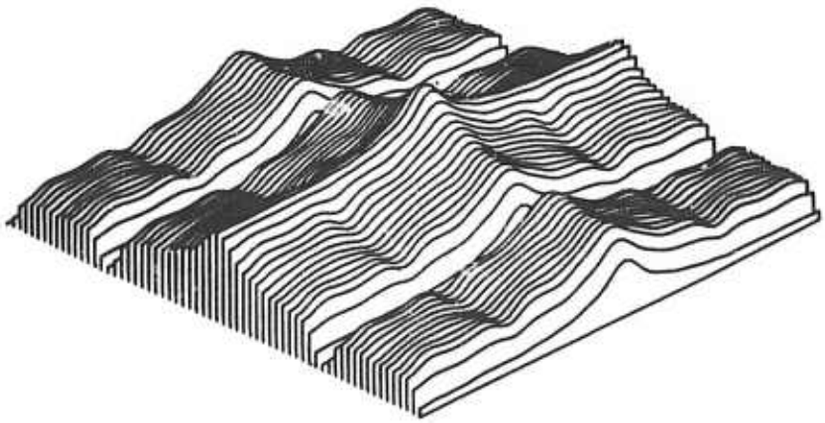


Figure 1.13 -- Equalization signal.

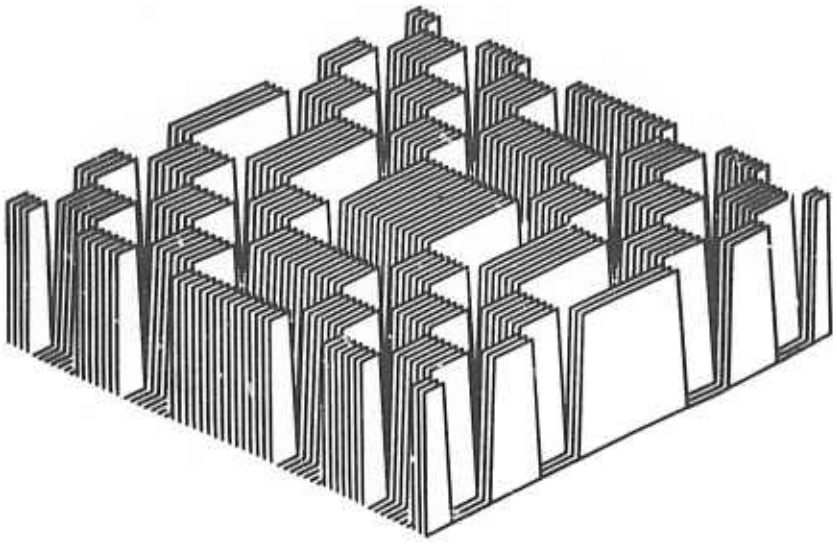


Figure 1.14 -- Phase of the artificial blur.

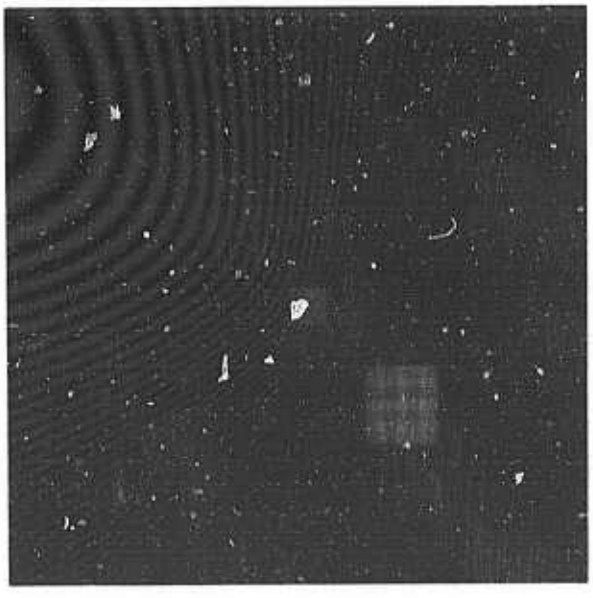


Figure 1.15 -- Image restored by the method of Figure 1.1.

Figure 1.16 -- A roadside sign blurred by an out of focus lens.

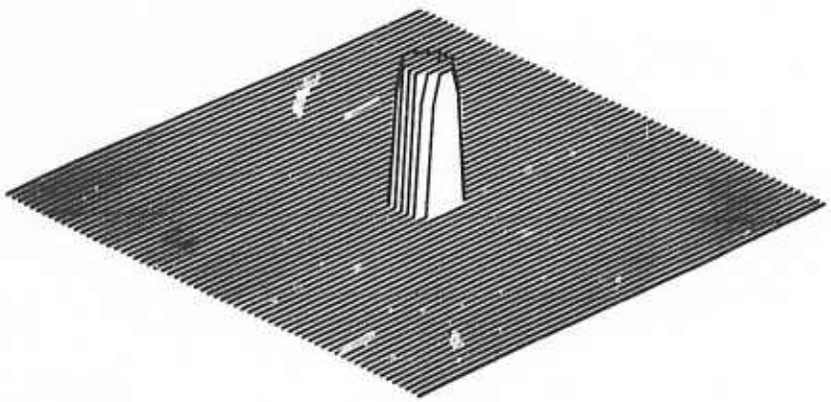


Figure 1.17 -- Stylized impulse response corresponding to the blur of Figure 1.15.

Figure 1.18 -- Restored image corresponding to Figure 1.16.

Figure 1.19 -- Sign blurred by camera motion.

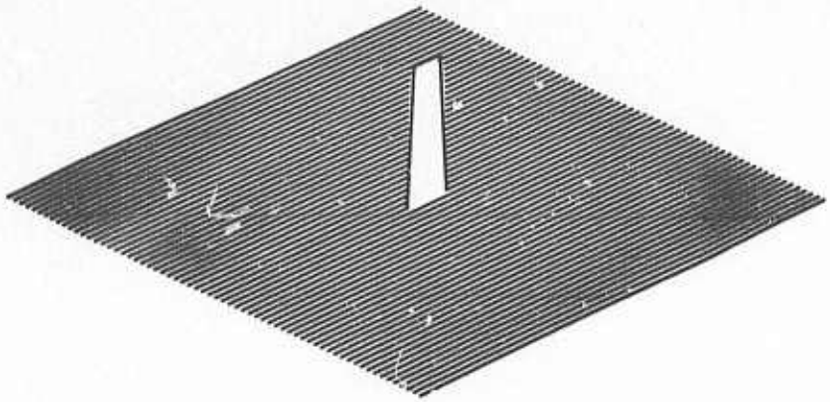


Figure 1.20 -- Motion blur impulse response. This figure represents a function that is zero except along a line in the center where it is a constant.

Figure 1.21 -- Restored image by the method of Figure 1.1. Note the effects on the right and the left edges of the images due to the interpolation process.

Figure 1.22 -- Sharp image of roadside sign.

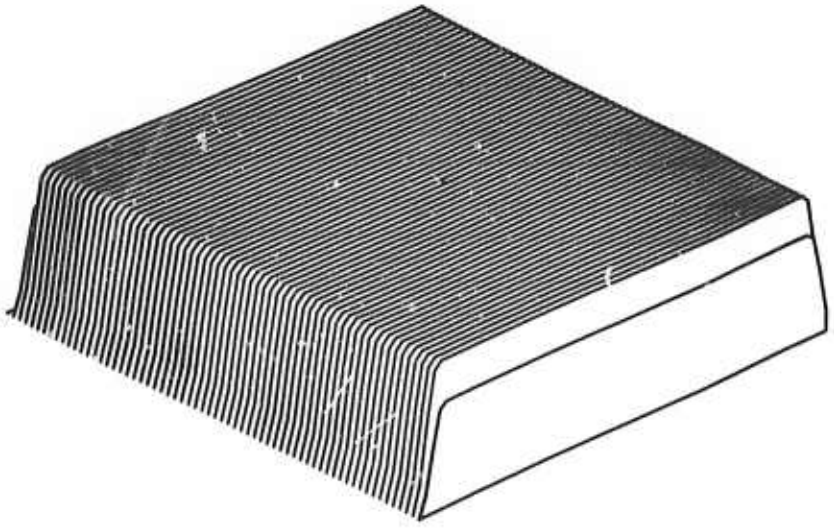


Figure 1.23 -- Interpolation function used to make the blurred image of Figure 1.16 have a constant value at the boundary.

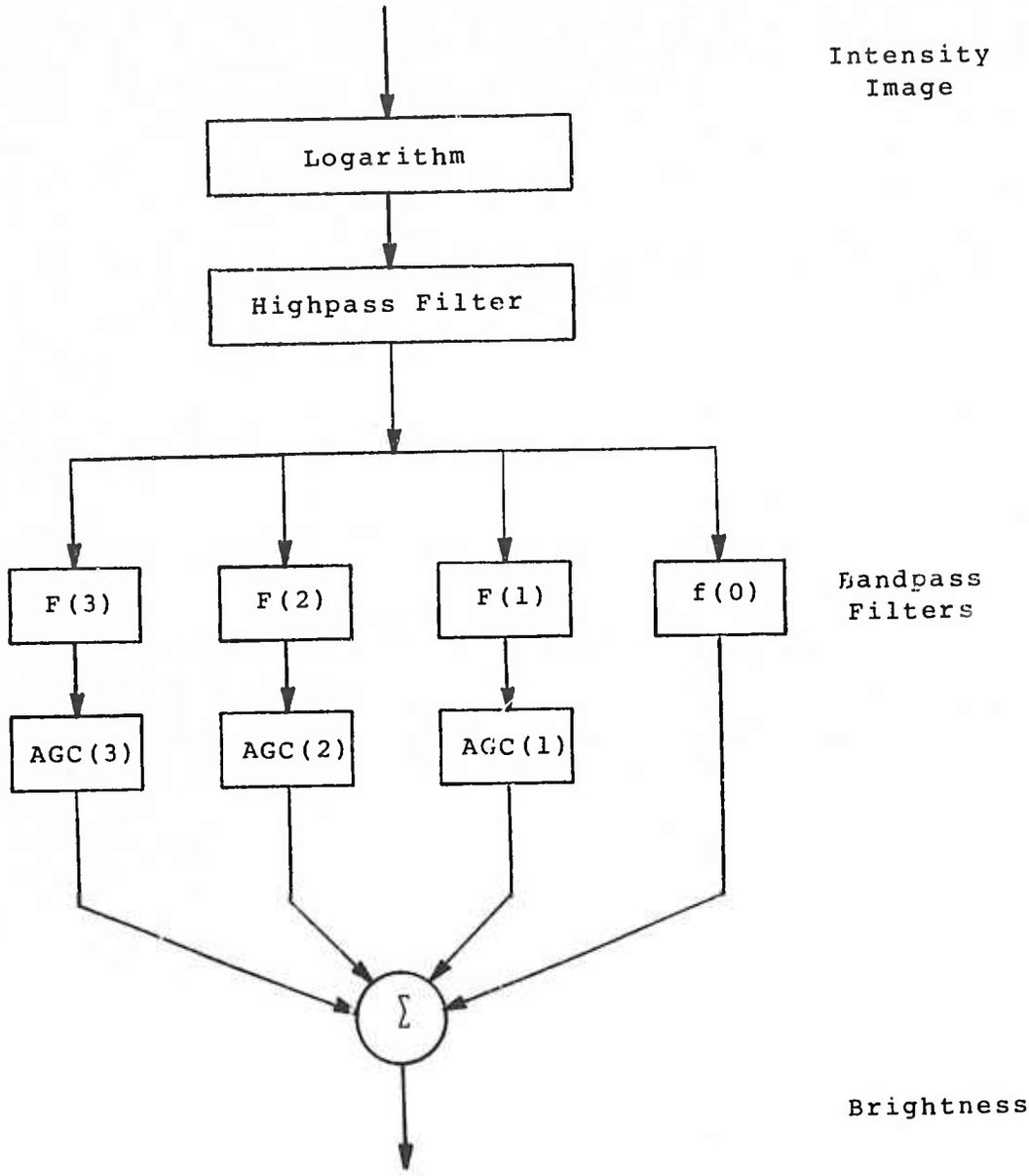


Figure 1.24 -- The frequency selective model of vision separates an image into channels based on spatial frequency content and does independent processing on each channel.

Section 2

The Modified Retinex Model

Brent Baxter

2.1 Introduction

An alternative to the multiplicative model which explains the visual system's ability to reject illumination effects, involves a kind of edge detector mechanism first proposed by Land [4,5]. Land's model correctly predicts the Cornsweet illusion but does not readily lend itself to image processing because it is not suitable for use on two dimensional sampled functions. A modification described in connection with the color constancy experiment overcomes this difficulty.

2.2 The Retinex

The term retinex was intended by Land to convey the impression of a neural network in the retina, optic nerve, and visual cortex for extracting lightness * information from images. To do this, ratios of light intensity are formed at adjacent points along a closed path and if a ratio is different from unity by only a small amount, it is set equal to unity. To extract lightness information from an image, the ratios are multiplied together around the path. The region crossed by the path corresponding to the largest accumulated product is assumed to be white and is given an arbitrary value against which the

---

* Lightness is Land's term for the psychophysical correlate of reflectance.

lightness at other points is compared. Necessary and sufficient conditions for lightness information computed in this way to be independent of the path are, that the image consist of patches each having uniform reflectance, and that the path cross the same reference, white, patch. Since in Land's formulation the path is arbitrary, computational complexity is introduced which makes the extraction of lightness quite difficult. A modification shown in Figure 2.1 overcomes this problem.

It is easy to see how such a system might reject information about the average level of illumination since all lightness values are scaled in a way that assigns reference white a predetermined value. Lightness values thus depend on ratios of reflected light rather than on the illumination level. Gradual changes in illumination across an image are also rejected by the retinex because light ratios within a patch will be close to unity and the threshold will set the ratio equal to unity. This eliminates any information about gradual illumination gradients. The color constancy experiment of Section 2.3 illustrates both of these properties. The new implementation of Land's retinex consists of taking the logarithm and applying a threshold to the magnitude of the gradient rather than applying a threshold to ratios around a closed path. This permits computations to be carried out on a rectangular grid in a systematic way. Integrating the gradient may be done to within an additive constant by an appropriate * line integral, and the constant is supplied in the process of assigning a predetermined lightness value to reference white. A color constancy experiment will be described next which

illustrates how the retinex can extract lightness information in the presence of illumination gradients.

2.3 A Color Constancy Experiment

Our interest in this experiment stems from the ability of visual system to perceive colored objects in correct color relationship to each other even under widely varying conditions of illumination. For example, a blue necktie looks nearly the same indoors under yellowish incadescent lighting as it does outdoors under much bluer natural light. This ability is known as color constancy. The retinex allows for this type of change in illumination and it can also adjust for gradual shifts in illumination hue across an image.

To test these ideas, a board covered with patches of colored paper was photographed in light from three slide projectors, each one fitted with either a red, green, or blue filter. The red projector, located upward and to the left of the board, supplied about twice as much light to the upper left corner as it did to the lower right corner, the green projector was positioned to supply about twice as much light to the lower right as to the upper left, and the blue projector illuminated the board evenly. These lighting conditions resulted in the pronounced red-green shift of Figure 2.3. Note how one of the white patches of Figure 2.3 appears pink and the other seems green. The visual system does not correct an illumination shift

---

* In the color constancy experiment an average of integrals taken over many paths of integration was used to minimize the effects of noise. See Wylie [11] for a discussion of line integration.

of this type on a reflection print as well as it would on a projected image in a darkened room. The reason is that clues from the book, paper etc. tend to inhibit the process. The red, green and blue components of Figure 2.4 were digitized separately and each one was passed through the retinex. In Figure 2.5 the lightness information is displayed. Note its similarity to Figure 2.2.

Natural scenes often consist of highly textured areas rather than the cartoon style patches prepared for this experiment. The retinex does not work well on texture and this is a serious defect. The reason is that individual patches in a textured image may be small enough to occupy a single picture element, and for the threshold to work properly, edges located by the gradient must not be too close together.

2.4 The Cornsweet Illusion

Figure 2.6 depends for its effect on an abrupt change in reflectance next to a gradual one. The abrupt change seems to be preserved by the visual system and the gradual one is attenuated. This property of the visual system has been taken advantage of for many years by artists and draftsmen in order to make a region in a drawing seem lighter or darker that it would otherwise. Ratliff [6] gives an interesting review of this practice. By a coincidence, this is the only illusion of the collection which is predicted correctly by the retinex and ignored by the frequency selective model. Figure 2.7 contains plots showing how lightness information may be extracted from the images in Figure 2.6. Note that the gradual change in intensity

has been completely eliminated while the abrupt changes are preserved.

2.5 Summary

The retinex model proposes to explain how lightness information may be extracted from images by computing ratios of reflected light intensity at adjacent points and retaining ones near the boundaries of objects. Neurophysiological studies have failed to reveal an organization of the type proposed by Land, but edge information is known to be important to perception [12]. The color constancy experiment of Section 2.3 shows how the retinex is capable of removing illumination gradients and Figure 2.5 shows retinex's correct response to the stimulus which produces the Cornsweet illusion.

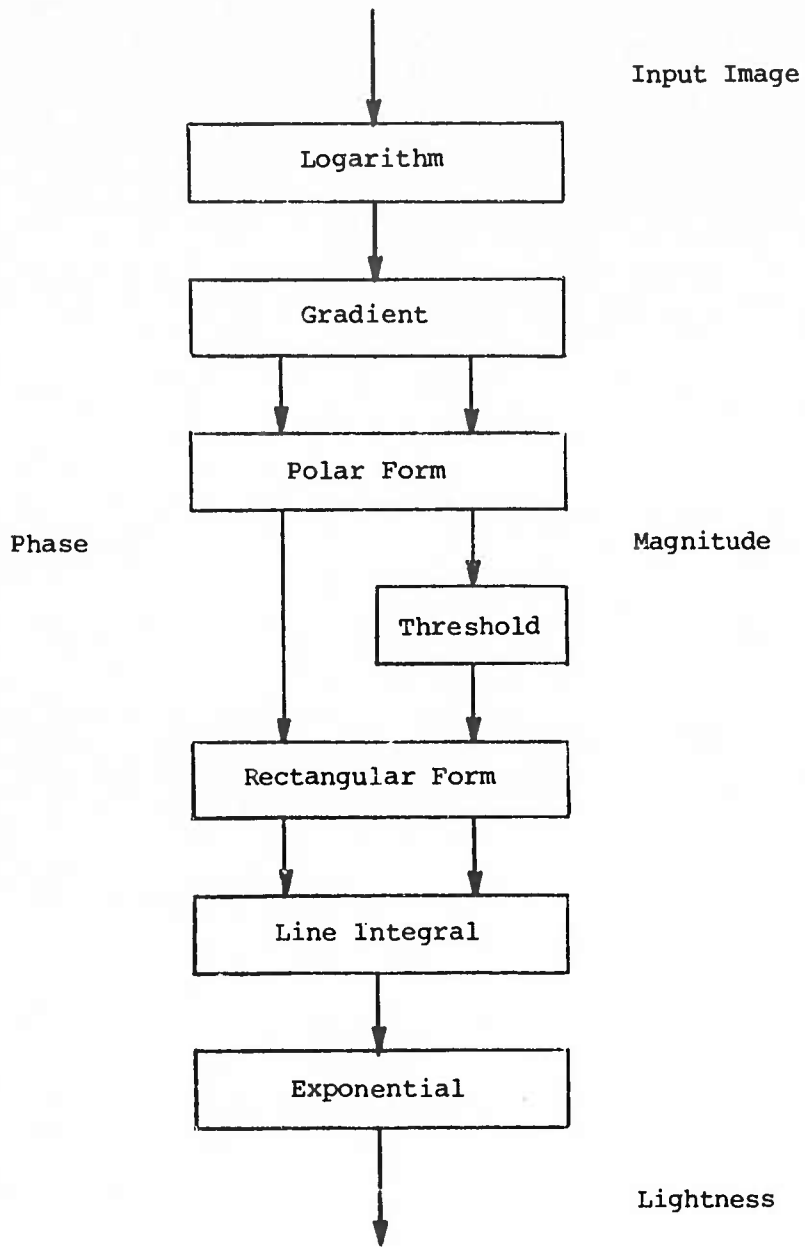


Figure 2.1 -- The retinex modified for image processing.

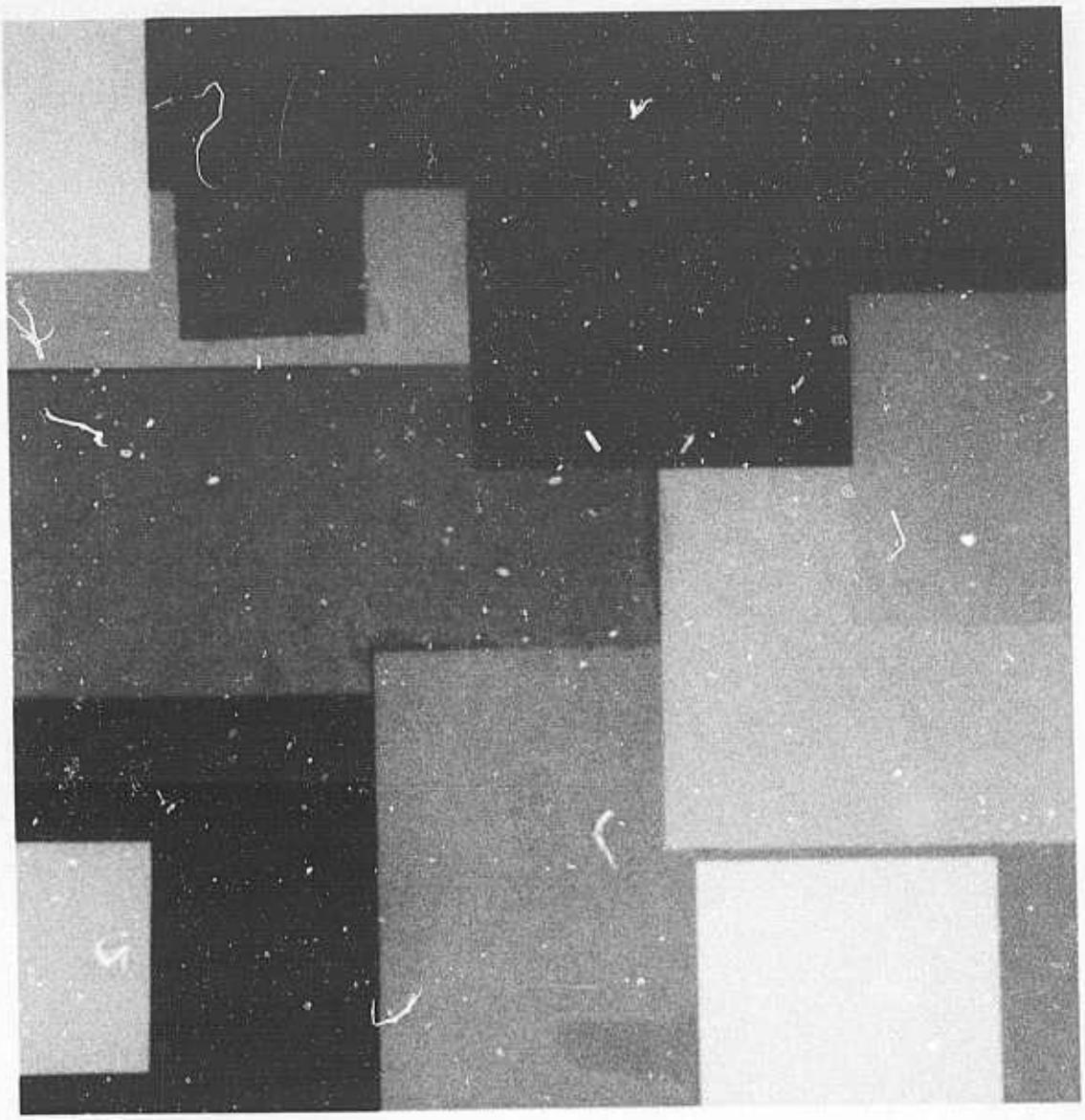


Figure 2.2 -- Test pattern photographed in white light.

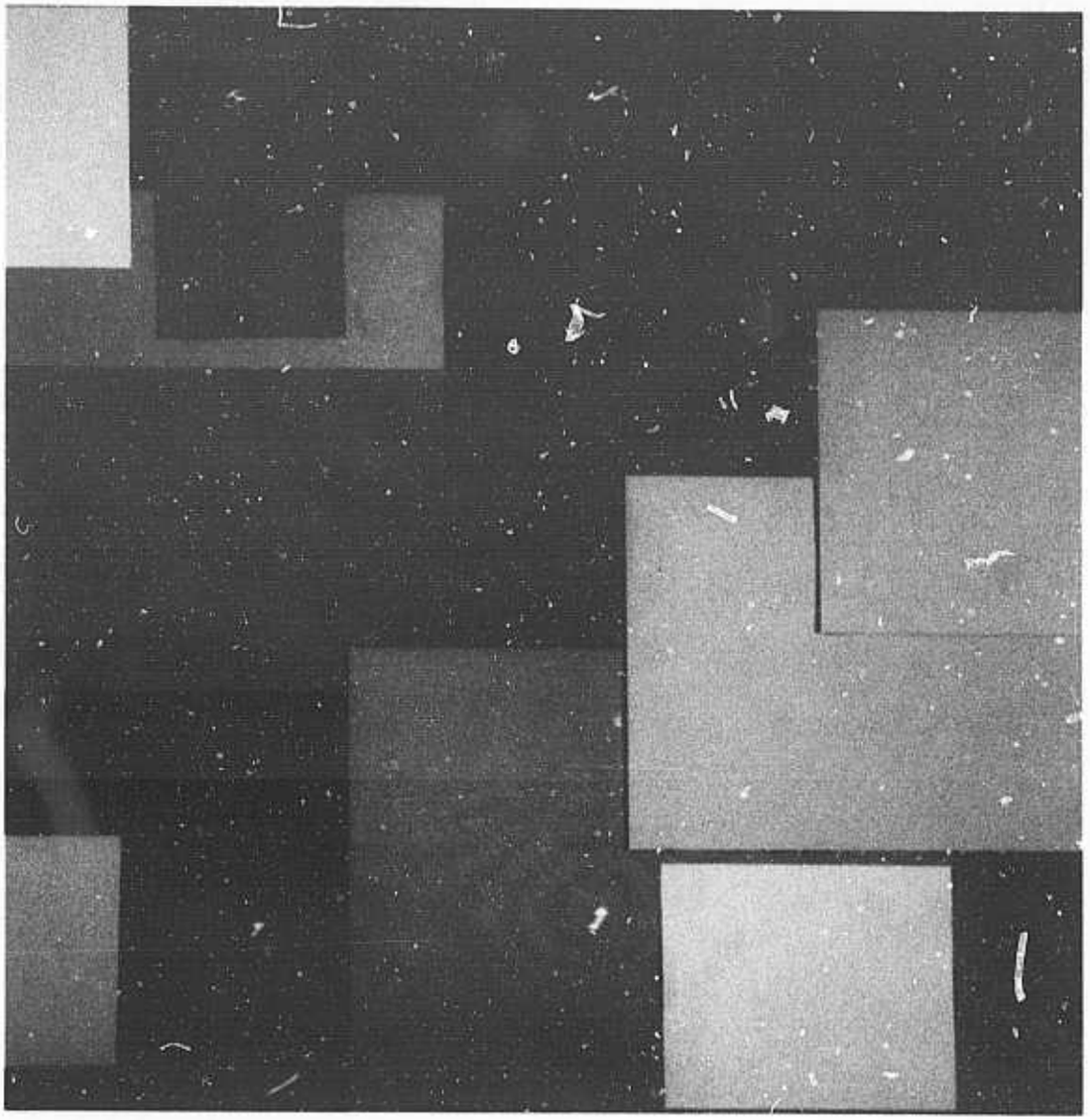


Figure 2.3 -- Test pattern photographed in colored light.

Figure 2.4 -- Digitized version of Figure 2.3.

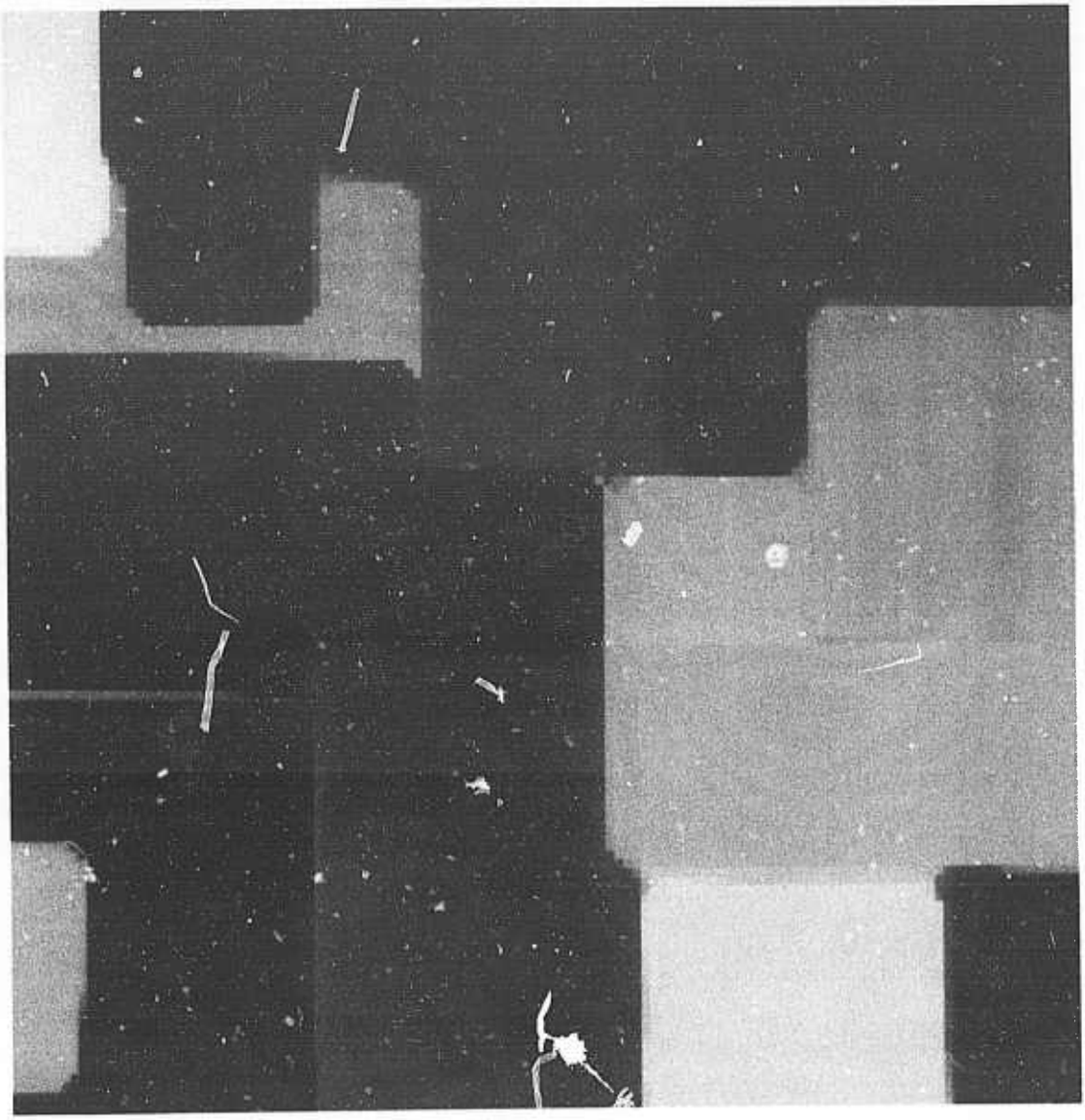
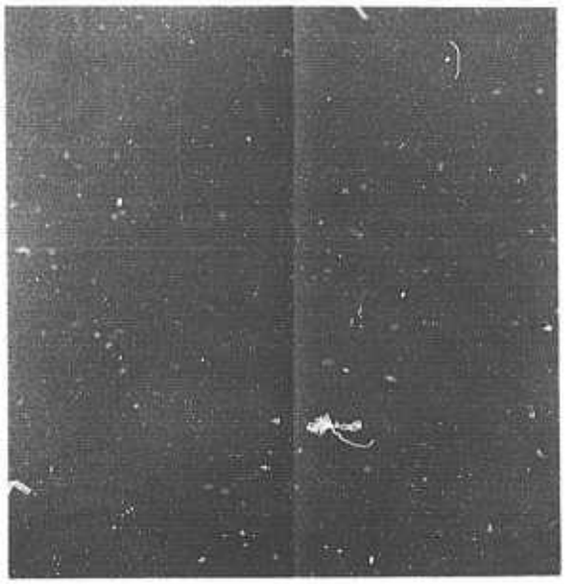


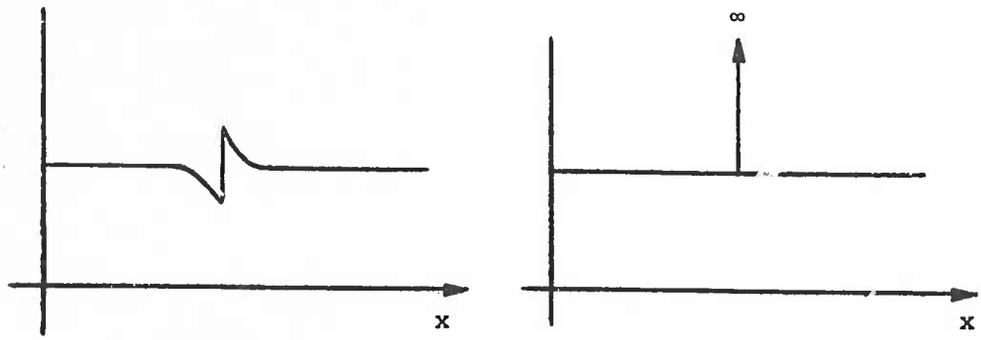
Figure 2.5 -- Color corrected pattern.

(a)



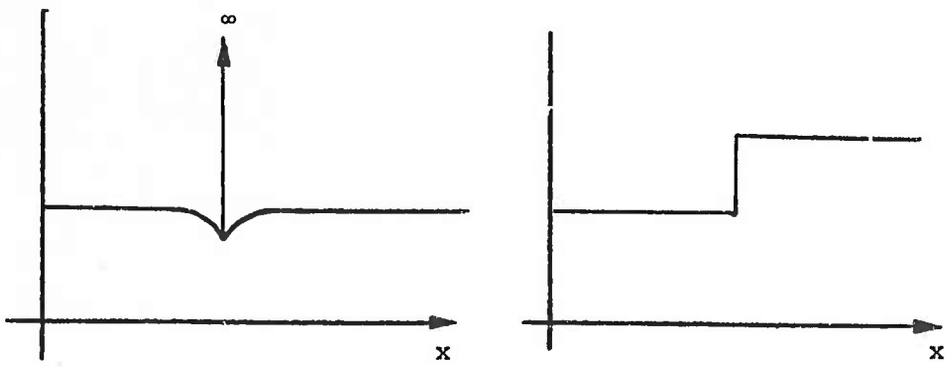
(b)

Figure 2.6 -- In these examples of the Cornsweet illusion, the left side (2.6a) appears darker than the right side (2.6b) even though they are an identical shade of gray at points a short distance from the edge.



(a) Intensity function of Figure (4a).

(c) Edge information.



(b) Gradient in the horizontal direction.

(d) Lightness information.

Figure 2.7 -- Extraction of lightness information from Figure 7.

Section 3

Conclusions
Brent Baxter

3.1 Review

The new frequency selective model of brightness perception described in the previous semiannual report, has been shown to give subjectively useful information greater impact where it is conveyed by weak texture. Properties similar to those of the multiplicative model are retained which help subdue undesirable illumination effects, and the spatial filters acting in concert with a bank of AGC elements help overcome a limited degree of blurring. The model is the basis for a significantly improved method for restoring blurred photographs that gives results free from echo-like artifacts common to earlier methods. Restorations of images blurred in the field by both an out of focus lens and by camera motion were presented to demonstrate the method. In Section 2, a new way of implementing the retinex was illustrated using a color constancy experiment as an example.

3.2 Speculation About Future Research

Physiological evidence is quite convincing that certain neurons in the visual cortex respond to different parts of the two dimensional Fourier transform of an image. This was illustrated by the adaptation experiment. The AGC elements are an attempt to model the adaptation part of the demonstration, however they were designed to make the

image enhancement work properly. This raises the question of whether they could be calibrated by an experiment similar to the one used by Baudelaire [7] in calibrating the multiplicative model. Another question which suggests itself from the appearance of Figures 3.1, 3.2 and 3.3 is whether this enhancement process could be used to make images more immune to the corrupting effects of coding. The multiplicative model has been used successfully for this purpose [8,9] and the AGC processing should bring about a further improvement. This would be a step forward with regard to finding a more accurate measure of image distortion.

Many questions about the image restoration process remain unanswered. One of them is the relationship between the cutoff frequency of the filter (Figure 1.11) and the quality of the restoration. Circularly symmetric filters were used in the restorations reported here but it seems possible that improved results would be obtained if the symmetry depended on the type of blur (motion, out of focus etc.). Also the preparation of equalization signals should be investigated further.

Figure 3.1 -- Natural scene containing variable strength texture.

Figure 3.2 -- Enhancement of Figure 3.1 based on the multiplicative model. The apple blossoms, which were quite prominent in the original (Figure 3.1), are further accentuated by the enhancement operation.

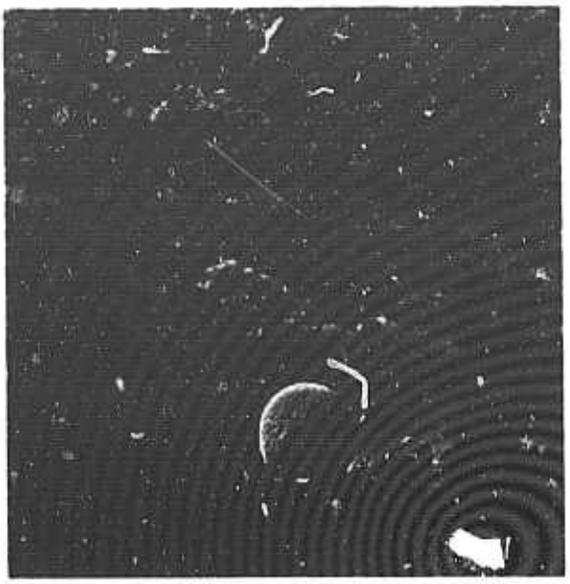


Figure 3.3 -- Image enhancement based on the frequency selective model results in a kind of texture equalization where weak detail in the dome and sky are strongly accentuated relative to corresponding parts of Figure 3.1 while the apple blossoms are given a softer rendition.

References for Sections 1, 2 and 3

[1] A.V. Oppenheim, R.W. Schafer and T.G. Stockham, Jr., "Non-Linear Filtering of Multiplied and Convolved Signals," Proceedings of the IEEE 56, (August 1968): pp 1264-1291.

[2] T.M. Cannon, "Digital Image Deblurring by Nonlinear Homomorphic Filtering," Ph.D Dissertation, University of Utah (1974).

[3] E.R. Cole, "The Removal of Unknown Blurs by Homomorphic Filtering" Ph.D Dissertation, University of Utah (1973).

[4] E.H. Land, "The Retinex," American Scientist 62 (1964): pp 247-264.

[5] E.H. Land and J.J. McCann, "Lightness and the Retinex Theory," Journal of the Optical Society of America 61 No. 1, (January 1971): pp 1-11.

[6] F. Ratliff, Mach Bands: Qualitative Studies on Neural Networks in the Retina, (San Francisco: Holden-Day, (1965): p 273.

[7] P. Colas-Baudelaire, "Digital Picture Processing and Psychophysics: a Study in Brightness Perception," Ph.D dissertation, University of Utah, (1973).

[8] J.L. Mannos and D.J. Sakrison, "The Effects of a Visual Fidelity Criterion on the Encoding of Images," IEEE Transactions on Information Theory IT-20 No.4, (July 1974): pp 525-536.

[9] R. Rom "Image Transmission and Coding Based on Human Vision," Ph.D dissertation, University of Utah (1975).

[10] T.G. Stockham Jr., "The Application of Generalized Linearity to Automatic Gain Control," IEEE Transactions on Audio and Electroacoustics AU-16, (June 1968): pp 267-270.

[11] C.R. Wylie, Jr. Advanced Engineering Mathematics (New York: McGraw-Hill, 1966): p 582.

[12] D.N. Graham, "Image Transmission by Two-Dimensional Contour Coding," Proceedings of the IEEE 55 No. 3 (March 1967): pp 336-346.

APPENDIX A

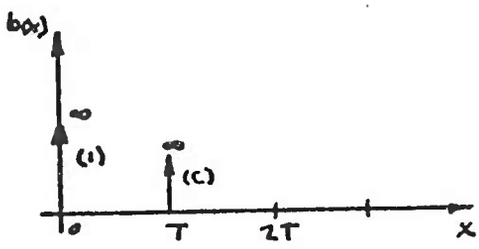
Notes on the impulse response length of convolutional inverses for space limited blurs.

Deconvolution is the process of separating signals that have been combined by convolution and it is ordinarily done with the aid of the Fourier transform.

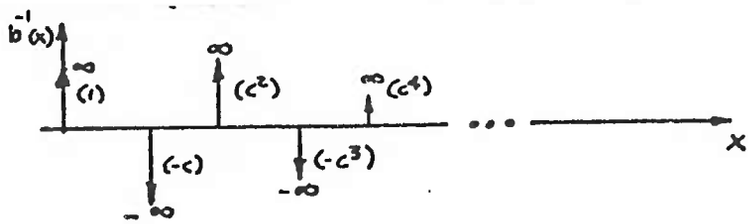
$$a = (a \circledast b) \circledast b^{-1} \xleftrightarrow{FT} (A * B)1/B$$

Suppose a signal is added to a delayed and scaled copy of itself. A convolutional representation of this situation is,

$$a \circledast b = a(t) + c*a(t - T)$$



The convolutional inverse of b can be found by inspection and is an infinite train of impulses.



Note that for c=1 the impulses do not tend toward zero for

large X.

A motion blur impluse response has a sampled representation much like the case for c=1 and its discrete convolutional inveree is of infinite extent. Truncating euch a sequence will introduce errors in the deconvolution.

APPENDIX B

Notes on truncation of discrete impulse response filters

Suppose two sequences, one of which is zero outside a finite interval, are to be convolved aperiodically.

$$g(j) = \sum_{i=-\infty}^{\infty} f(i) \ast h(j-i)$$

where, $f(i) = 0$ for $i<k$ and $i>l$, $k<l$

If $h(j)$ is nonzero $\ast$ over the infinite interval $\infty < j < \infty$ the sum above must be taken over all i and $g(j)$ will also be nonzero over the infinite interval.

Computational difficulties implied by the infinite sum may be avoided if $g(j)$ is only required over part of its domain as in image processing where the output image is often made equal in size to the input image. If $g(j)$ must be computed only for $m<=j<=n$, then

$$g(j) = \sum_{i=k}^{l} f(i) \ast h(j-i) \qquad m<=j<=n$$

$l-k+1$ terms are required in the sum, the sum must be evaluated $n-m+1$ times and $h(j)$ must be available for $m-l<=j<=n-k$.

If the lengths of input and output sequences are l-k+1 and n-m+1, the impulse response must be m-l+n-k+1 points long. For example, if input and output sequences are 256 points long, 511 points of the infinite impulse response, n(j), will be required. Similar arguments apply in two dimensions.

---

* By the expression nonzero over an interval it is meant that elements of the sequence may be nonzero within the interval but they are zero outside the interval.

APPENDIX C

Circular deconvolution of periodic sequences by filtering their log spectra

Periodic sequences combined by circular convolution may be separated using the complex logarithm * and the convolution property of the discrete Fourier transform (DFT). If a,b are sequences having the same period and A,B are their discrete Fourier transforms, then

$$a = (a \circledast b) \circledast b^{-1} \overset{FT}{\longleftrightarrow} (A * B)/B \overset{Log}{\text{-----}} (\overline{A} + \overline{B}) - \overline{B} = \overline{A}$$

The square overbracket indicates the complex logarithm.

B may be removed by direct subtraction if it is known exactly or by linear filtering if its DFT is disjoint from the DFT of A.

APPENDIX D

An Image Processing Language (IPL)

Experimente of the type described herein were greatly facilitated by the development of a programming language [29] whose data typee are images and whose primitives operate on images. Typical IPL primitives allow manipulation of imagee through addition, the discrete Fourier transform and magnitude-phase computations to name but a few. As a further experimental convenience, these operations may be invoked separately as a command language by typing interactively on a computer terminal or they may be combined into an IPL program and interpreted by the language processor. The convenience provided by thie approach makes it feasible to explore many more poeeibilities than if each experiment required that a separate program be written, compiled and debugged using traditional editore, compilers and debugging aids.

Section 4

Color Image Processing in the Context of a Three-Dimensional Homomorphic Model

Olivier Faugeras

We have been working for the last six months on a model of human vision which extends Stockham's model [13]* to include color phenomena and can also include the new ideas involving the existence of separate frequency channels in our brightness vision system. Although the old [12,2] and new [3] results concerning black and white vision have been constantly used in our research, our major effort has been directed toward an understanding of problems related to color. The model which takes into account the most recent knowledge about the physiology of color vision is, in its oldest version, a three-dimensional homomorphic system and is thus built on ideas this group has been familiar with for a long time and have been shown to be successful and fruitful.

To get straight on notations, let us review some related mathematics: a color image can be represented as a function I of three variables x, y, λ where x and y are the spatial coordinates and λ is the wavelength of the light reflected from the print.

To explain the results of trichromatic matching experiments, it has been hypothesized that the human retina has three types of cones with three different absorption curves. This hypothesis has recently been confirmed by reflexion densitometry measurements in the living eye of normal man [1] and by absorption spectrum

---

* See references at the end of this section.

measurements of single cones in excised retinas from man and monkey [9].

The effect of this cone absorption can be modelled by the following equations:

$$J_i(x,y) = \int I(x,y,\lambda)a_i(\lambda)d\lambda \qquad i = 1,2,3 \tag{1}$$

where the integral is taken over the visible spectrum and where $a_i(\lambda)$ $(i = 1,2,3)$ is the absorption spectrum of the ith type of cone.

The next fact to be considered in the model is that the total cone response is not linear but a monotonically increasing convex function which can very well be approximated by a logarithm function [4] [5]:

$$D_i(x,y) = \mathrm{Log}[J_i(x,y)] \qquad i = 1,2,3 \tag{2}$$

The most recent physiological studies [7] [5] have also shown that after those two stages lateral inhibition was present between cones, fact that we can model by the cascade of two systems:

--one amnesic linear system defined by:

$$\begin{cases} G_1(x,y) = \alpha * D_1(x,y) \\ G_2(x,y) = \beta * (d_2(x,y) - d_1(x,y)) \\ G_3(x,y) = \gamma * (D_3(x,y) - D_1(x,y)) \end{cases} \tag{3}$$

where α, β, γ are three carefully chosen constant numbers (Frei).

--One linear space invariant system with memory, defined by three impulse responses $h_1(x,y)$, $h_2(x,y)$, $h_3(x,y)$ or equivalently by three frequency responses $H_1(f_1,f_2)$, $H_2(f_1,f_2)$, $H_3(f_1,f_2)$:

$$T_i(x,y) = G_i * h_i(x,y) \qquad i = 1,2,3 \tag{4}$$

* denoting two-dimensional convolution

At this point a block diagram of the total model might help:

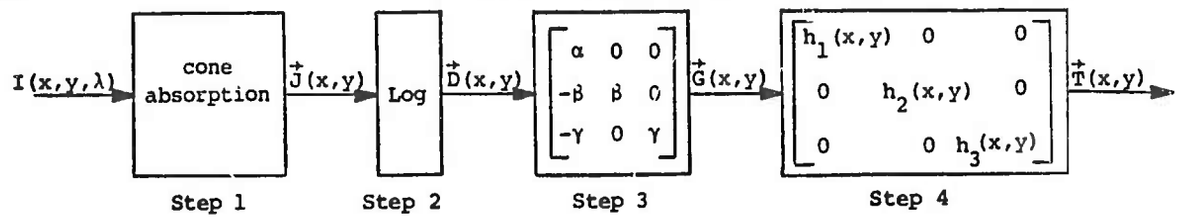


Figure 1

The vector $T(x,y)$ can be considered as the representation in some three-dimensional "perceptual space" of the original image $I(x,y,\lambda)$.

It must be pointed out that all steps but one, namely the first, in this model are invertible. This non-invertibility is not really going to hamper us since it has been known since Helmholtz that any image can be represented in terms of only three primaries (red, green, blue for example). This indicates that although the first transform is destroying almost all wavelength information, this fact is of no consequence at all. Actually, for the purpose of our work, we might as well entirely ignore it and consider that the images we are working with are defined as three images or equivalently as a vector $\vec{I}(x,y)$, the three components of which

correspond to the red, green and blue contents of the original picture.

The cone absorption transform now becomes a simple amnesic linear transform corresponding to a change of coordinate system. The change is from the one used to obtain the three separation prints to the one defined by the three cones absorption curves, thus making the whole system entirely invertible.

Before going into the work we have been doing with this model, let us review a few more interesting facts concerning its relationship to previous and recent models for black and white vision that have been successfully used in this group: It can be shown that one cone absorption curve is very closely approximated by the CIE relative luminous efficiency function and thus it is built into the model that the brightness information is present in only one channel (we chose the first one or the first coordinates) of every vector; this brings up the fact that in the form of Figure 4.1 and for a black and white image (for which all three components of $\vec{J}(x,y)$ are equal) $\vec{G}(x,y)$ is of the form $[G_1(x,y), \phi, \phi]^T$ where $G_1(x,y)$ is the brightness information. Thus, the frequency response $H_1(f_1, f_2)$ is known to us and can be taken as the one used by Stockham [13] or as the one measured by Baudelaire [2].

In this sense, this model can be considered as an extension of Stockham's. Also, since all brightness information is contained in Channel 1, it can readily be seen that more recent results concerning the existence of separate frequency channels [3] are

easily included by just acting on Channel 1 after Step 3.

We are now free to concentrate our attention on Channels 2 and 3 which correspond to the chromatic information and especially on the two frequency responses $H_2(f_1, f_2)$ and $H_3(f_1, f_2)$. Very little work has been done to measure those modulation transfer functions and it is one of the purposes of our study to contribute to this measurement and show that some phenomena related to color illusions that have puzzled people for long (color constancy, color contrast) can be explained by this model. Another purpose of this work is to show that processing similar to the one performed by Stockham on black and white images can also be done in color. Finally, a third purpose is to study the effects of a visual fidelity criterion defined by the model for the encoding of color images.

It must be pointed out that all of these purposes are closely related. To give an example: the human visual system can easily discard chromatic illumination in a very large range of intensities (color constancy effect as shown by experiments by Land [8]), this fact gives us sketchy information on the shape of H_2 and H_3 near the origin; this information in turn can be used both to perform enhancements and also to preprocess prior to coding. We now proceed to describe our accomplishments.

4.1 Color Image Processing

Let us first review our work with the processing of color images. The main idea is two-fold: in the processing of a color

image we might want to do two things:

--Increase the saturation of objects present in the scene. Since objects on a picture tend to be small, this implies that our frequency responses H_2 and H_3 should have large values at high frequencies.

--Get rid of any chromatic illumination that might be caused either by a small error in the color balance during the printing process or by the actual presence of a strong chromatic illumination. This implies that H_2 and H_3 should have low values (less than 1) near the origin.

Using those ideas, an experiment has been devised to enhance color images. Real-life color pictures have been scanned into the computer using the separation print method:

A color negative is used to produce three black and white prints on paper exposed through three different filters (Red 92, Green 99, Blue 478), each print being afterwards separately scanned in.

The results are shown on Plates I and II: the "original" reproduced from the original bits stored on disk, and the "chromatic enhancement." This latter is obtained by processing every image using the model of Figure 4.1 with:

$h_1(x,y) = \delta(x,y)$ (no brightness processing)

$H_2(f_1, f_2)$ and $H_3(f_1, f_2)$ have been taken to be circularly symmetric with the cross-sections shown on Figure 4.2.

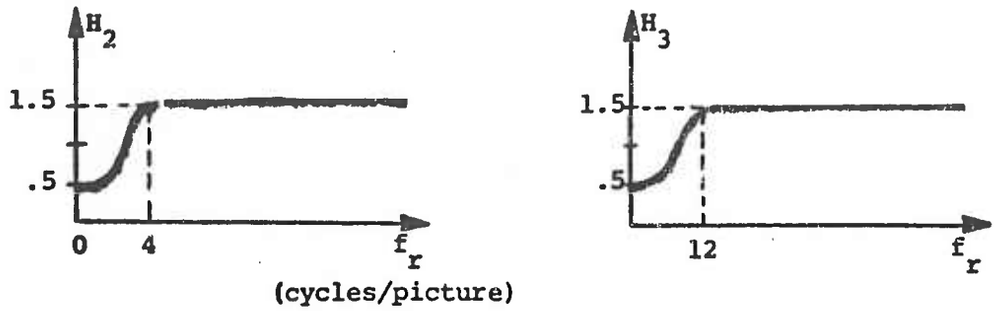


Figure 4.2

Also shown are the same image processed again by the previous H_2 and H_3 but by taking also $H_1(f_1, f_2)$ cross-section as shown in Figure 4.3.

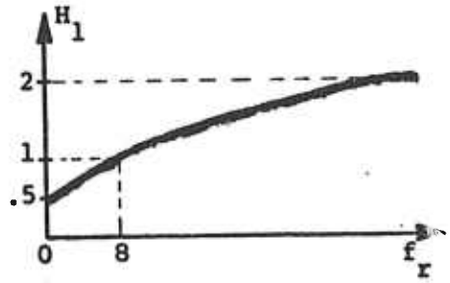


Figure 4.3

This last processing intends to show that results obtained previously by Stockham in his enhancement work with black and white images can be successfully extended if applied to the brightness channel of this color model which was by no means obvious beforehand. Results are shown on Plate III.

This second processing may be called "total homomorphic enhancement."

4.2 The Use of a Distortion Measure in the Encoding of Color Images

We also said before that we wanted to investigate the possibilities offered by the model to define a distortion measure between still color images

Let $\vec{I}(x,y)$ be our original image that we want to transmit over some noisy channel and $\vec{\tilde{I}}(x,y)$ the received image. We would like to be able to define a distance or distortion measure between these two images such that $d(\vec{I},\vec{\tilde{I}})$ is in agreement with subjective evaluation. The reason why we are interested in such a distortion measure is the following: Shannon's rate-distortion function [12] provides a useful lower bound against which to compare the rate-versus-distortion performance of practical encoding-transmission systems by the following: The distance d being defined, the performance of the system is measured by the average distortion:

$$d^* = E\{d(\vec{I},\vec{\tilde{I}})\} \qquad (5)$$

where the expected value is taken over the ensemble of images of interest. Shannon's rate-distortion function $R(d^*)$ is a lower bound of the transmission rate required to achieve average distortion d^*; moreover, Shannon's coding theorem also states that one can design a code with rate only negligibly greater than $R(d^*)$ which achieves average distortion d^*. This function $R(d^*)$ thus exactly specifies the minimum achievable transmission rate R' required to transmit an

image with average distortion level d^* and provides an absolute yardstick against which to compare the performance of any practical system.

To date, this potential value has not been realized for image transmission for several reasons. One reason is that there does not currently exist any tractable mathematical models for an image source. A second reason is the difficulty of calculating the rate-distortion function for other than Gaussian sources and square-error distortion measures.

However, it is generally recognized that the prime reason that rate-distortion theory is not applicable is that a distortion measure in agreement with subjective evaluation of image quality is not known. Since the model we are using embodies the most recent knowledge about black and white and color vision, it is very appealing to use it to define a class of distortion measures in order to calibrate and test it further. It would then be possible to compare different distortion measures in the class by simulating the encoding of a fixed image at a fixed rate under different distortion measures and subjectively judging the quality of the encoded images.

If, for a variety of source images and rates of interest, many subjects uniformly pick images encoded under the same distortion measure as appearing best, then clearly that measure is the most appropriate in the class to use for evaluating transmitted images.

Let us now describe the class of distortion measures defined by the model: remember we assume that a color imagee ie a three-dimensional vector $\vec{I}(x,y)$ where x and y are the epatial coordinates and that the model is a homomorphic system mapping $\vec{I}(x,y)$ to $\vec{T}(x,y)$. Now given two images $\vec{I}(x,y)$ and $\vec{\tilde{I}}(x,y)$, we can define the distance between them as a set of three real poeitive numbers α_1 α_2 α_3 such that:

$$\alpha_1 = \iint [T_1(x,y) - \tilde{T}_i(x,y)]^2 \, dx\,dy \qquad i=1,2,3 \tag{6}$$

(Notice that this is not a distance in the normal mathematical sense.) If one assumes that for all images considered, only the second-order statistics (mean and correlation function) are known then for distortion measures of the class we are considering, the Gaussian distribution is the worst in the class of all probability distributions of a random field with given mean and correlation function [11] [12]. But this would not be of great interest without the following result that the optimum code for the Gaussian source which yields average distortion d^* is robust, i.e., it yields average distortion less than or equal to d^* for any source in the class [10].

This eolves the problem of not knowing the statistics of the source. It allows us to compute the rate-distortion function for our class of source distributions as the rate-distortion function of the Gaussian source and simulate the optimum encoding for the Gaussian source.

It was noted that the homomorphic model defined a class of distortion measures. By this, it is meant that some parameters defining the model were going to be made vary. Let us now be more specific about this point. So far we have only investigated the effect of varying the two frequency responses H_2 and H_3. We did not investigate variations of H since this has been done recently [11]. These results were closely correlated to these obtained by Stockham and Baudelaire in a different context [13] [2]. Similarly, we did not question any other part of the model like stage 3 [5], stage 2 [11], or stage 1 [6] [14], but this obviously could be done. Chromatic frequency responses have rarely been measured although some work has been done on this subject but never to our knowledge in a framework similar to ours. Just as for the brightness frequency response, the chromatic frequency responses are assumed to be circularly symmetric, but with the high frequency rolloff thought to occur much sooner. There is quite a lot of controversy about the existence of a low frequency rolloff similar to the one occuring for brightness.

Those two points are very important for different reasons: the early high frequency rolloff corresponds to the fact well known to television engineers that color information needs a much narrower bandwidth than brightness in order to be transmitted with satisfying accuracy. The low frequency rolloff would account for such phenomena as color constancy and simultaneous color contrast.

In order to test those ideas, an experiment has been designed

which can be described as follows: The optimum encoding and transmission of a color image $\vec{I}(x,y)$ has been simulated in the sense explained above at two different rates (.1 bits/pel and .05 bits/pel), and for different frequency responses H_2 and H_3 chosen from a set of 4 filters (see Figure 4.2) which peak at 1, 2, 4 and 6 cycles/degree. At both rates, the pictures looking the most like the original as far as color information is concerned are the one processed through the model where H_2 and H_3 are peaking at 1 or 2 cycles/degree. At this point, it might be useful to insist again on the fact that we are exclusively interested in color information so that in all these simulations the brightness information is left untouched. These results strongly suggest a very early high-frequency rolloff of the chromatic frequency responses, which is in strong agreement with other studies.

The filters used all have a low-frequency rolloff. Further studies are planned in order to explore the influence of rolloff characteristics. Low frequency rolloff was investigated first because of the experimental results described in Part 4.3. The "original" CAR-PORT is shown on Plate I The result of the simulation at .05 bits/pel with H_2 and H_3 peaking at 6 cycles/degree is shown on Plate IV, and 2 cycles/degree is shown on Plate V.

4.3 Color Illusions and Psychophysics

Finally, we also said that we were going to contribute to the measurement of chromatic modulation transfer functions. It has been shown by Baudelaire [2] that many brightness illusions such as

simultaneous contrast, Herman Grids, and Mach Bands can be explained by a low frequency alternation in H_1. The results of his study show that in the range 0-2 cycles/degree, which is the range of frequencies where brightness contrast effects are the most prominent, the following relation was approximately true:

$$H_1(3\underset{\sim}{f}) = 2H(\underset{\sim}{f}) \quad (\underset{\sim}{f} \text{ is the radial frequency}) \tag{7}$$

This relationship indicates a strong low frequency rolloff in H. We started to investigate whether effects similar to simultaneous brightness contrast were obtainable; the answer was found to be yes. These effects are called simultaneous color contrast. Compensation experiments conducted on the author have indicated that a relation similar to (7) seems to hold for H_2 and H_3. These results need to be made more precise by experimenting on several observers and standardizing the viewing conditions.

4.4 Future Directions

We essentially plan to go on with experiments related to topics 4.2 and 4.3 in order to get more precise and more complete results.

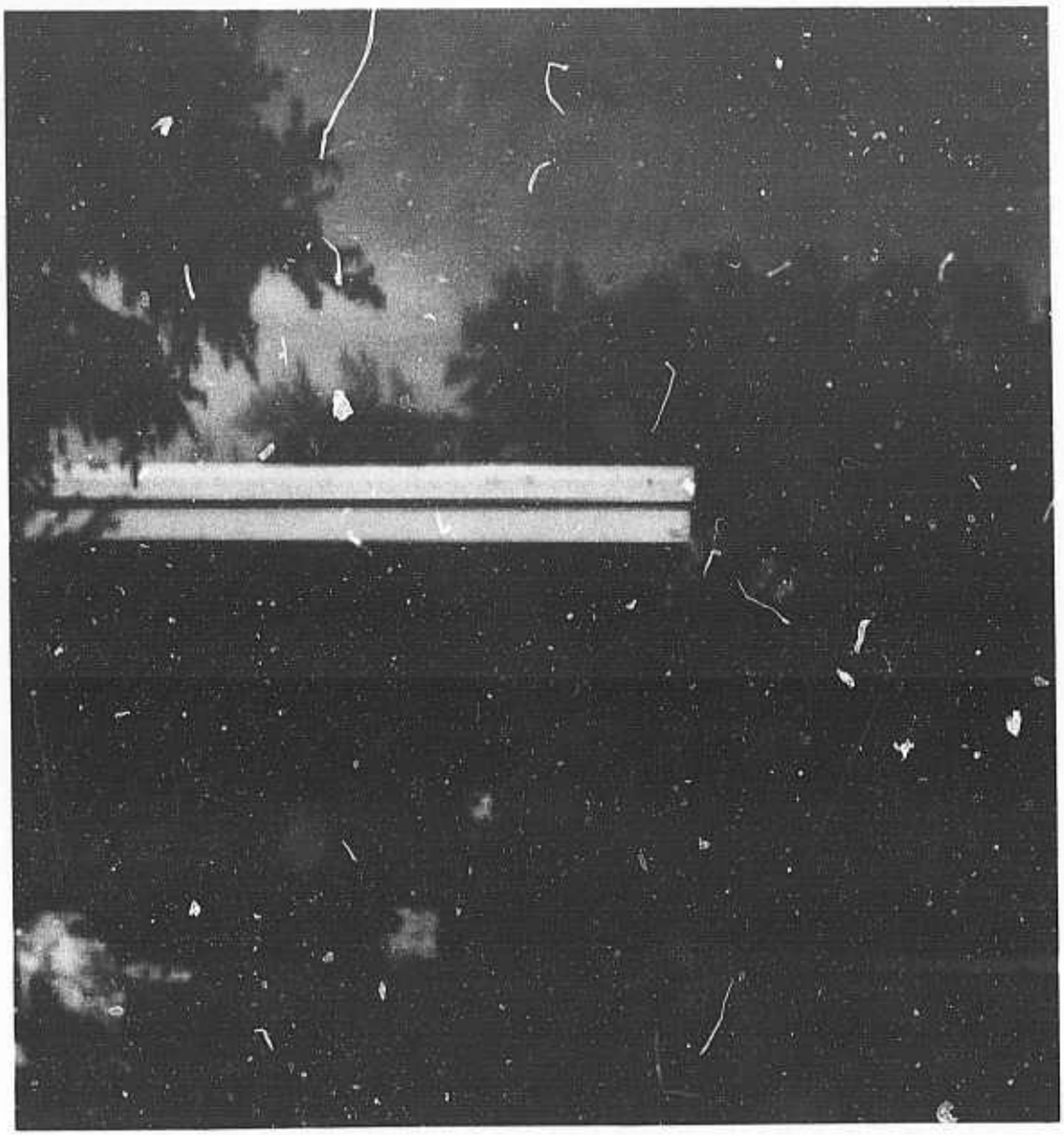


Plate I -- Original

Plate II -- Chromatic Enhancement

Plate III -- Brightness and Chromatic Enhancement

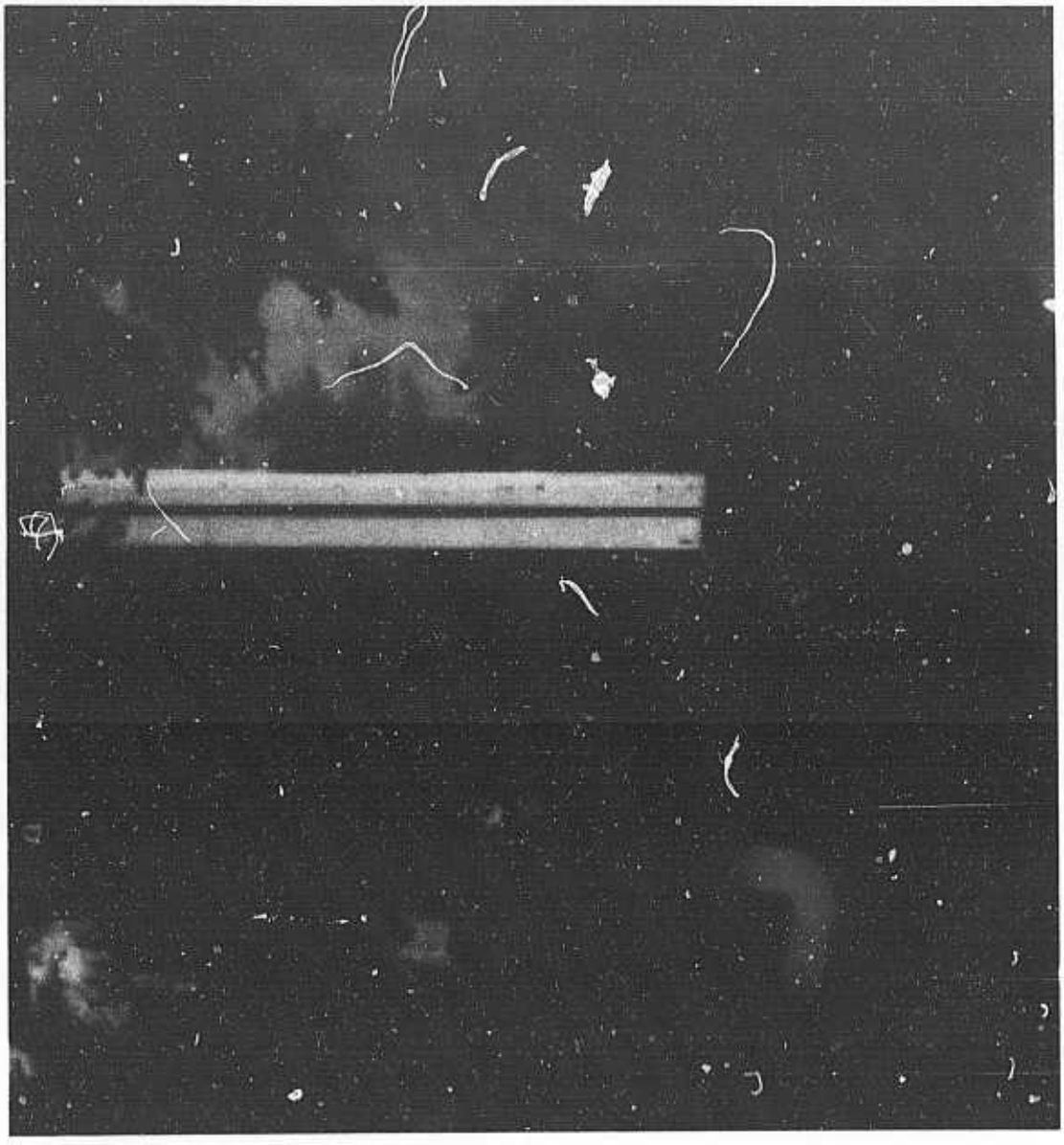


Plate IV -- Simulation of Encoding at .05 bits/pel
H2 and H3 Peaking at 6 cycles/degree

Plate V -- Simulation of Encoding at .05 bits/pel
H2 and H3 Peaking at 2 cycles/degree

References for Section 4

[1] Baker, H.D. and W.A.H. Rushton. "The Red-Sensitive Pigment in Normal Cones." Journal of Physiology 176 (1965), pp 56-72.

[2] Baudelaire, P. "Digital Picture Processing and Psychophysics: A Study of Brightness Perception." Ph.D. Thesis, University of Utah, Computer Science Department (1971), Salt Lake City, Utah.

[3] Baxter, B. "Image Processing in the Human Visual System." Ph.D. Thesis, University of Utah, Electrical Engineering Department (1975), Salt Lake City, Utah.

[4] Cornsweet, T.N. "Visual Perception." New York:Academic Press (1970).

[5] De Valois, R.L. "Central Mechanisms of Color Vision."

[6] Frei, W. "A Quantitative Model of Color Vision." University of Southern California Semi-Annual Technical Report (30 September 1971).

[7] Hubel, D.H. and T.N. Wiesel. "Receptive Fields and Functional Architecture of Monkey Striate Cortex." Journal of Physiology 195 (1968), pp. 215-243.

[8] Land, E.H. "The Retinex." American Scientist 52, No. 2 (1964), pp. 247-264.

[9] Marks, W.B.; Dobelle, W.H. and E.F. MacNichol. "Visual Pigments of Single Primate Cones." Science, N.Y. 143 (1964), pp. 1181-1183.

[10] Sakrison, D.J. "Notes on Analog Communication." Chapter 6. New York:Van Nostrand Reinhold (1970).

[11] Sakrison, D.J. and J.L. Mannos. "The Effects of a Visual Fidelity Criterion on the Encoding of Images." IEEE Transactions on Information Theory IT-20, No. 4 (July 1974).

[12] Shannon, C.E. "A Mathematical Theory of Communications." Bell System Technical Journal 17 (October 1948), pp. 623-656.

[13] Stockham, T.G., Jr. "Image Processing in the Context of a Visual Model." Proceedings of the IEEE 60, No. 7 (July 1972).

[14] Wyszecki, G. and W.S. Stiles. "Color Science." New York: John Wiley and Sons, Inc. (1967).

Section 5

Spectral Matchng Using LPC
Steven F. Boll

A technique for matching two digitized waveforms wae developed. Two waveforme are compared by using the predictor coefficients from one waveform and the data samples from the other to generate a mixed parameter error signal. Dividing the energy, of thie error signal by the minimum error energy, defines a ratio whose divergence from one measures their dieeimilarity.

This technique was applied to two areas: (1) speaker authentication; and (2) isolated word recognition. For the speaker verification experiment, the sentence "May we all learn a yellow lion roar." as spoken by ten speakers, was recorded on six occasions, three each day, spaced a week apart. On this data, a verification accuracy of 95% was obtained. In parallel with this effort, was the application of this waveform matching technique to the problem of isolated word recognition. A vocabulary of 107 worde recorded nine times was used. Necessary auxiliary algorithms were developed for word boundary detection, non-linear time regietration by dynamic programming, removal of redundant information and improved reference parameter sets through averaging. A recognition accuracy of 97% was obtained. The theoretical development of the waveform matching technique and detaile of the speaker verification experiment are described in University of Utah Computer Science Technical Memorandum 7500, May 6, 1975.

Section 6

Improving Synthetic Speech Quality of Low Bit Rate LPC Vocoder Systems

Steven F. Boll

Methods for improving low bit rate synthetic speech quality were developed both by modifying the existing LPC vocoder systems and by investigating new techniques of speech analysis. Synthetic speech obtained by linear prediction is minimum phase and when listened to on headphones exhibits an annoying buzzy quality. However, when heard in room over loudspeakers where the phase has been dispersed, the buzziness is less noticeable and the quality subjectively better. By modeling the room as a linear stationary system, the effect of room reverberation on synthetic speech heard over a headset can be approximated by measuring the room's impulse response as recorded by two microphones spaced the ear's distance apart, convolving the synthetic speech with each of the impulse response, and playing the resulting reverberated speech through each headset channel. Preliminary experiments demonstrate that the annoying qualities of LPC speech are noticeably reduced using this technique. Further refinements to this process of applying binaural reverberation are being considered.

A new technique of speech analysis called cepstral prediction is being investigated. By applying linear prediction to the cepstrum of the signal, a synthetic waveform can be generated whose spectrum matches the logarithm of the original spectrum rather than the spectrum itself. The advantages of log spectral matching using

cepstral prediction over standard LPC for generating high quality, low bit rate synthetic speech, are being investigated. Also being considered is the possibility of using this technique to analyze speech corrupted by both additive and convolutional noise, where it was demonstrated (Miller, 1974) that homomorphic processing suppresses noise in acoustic recordings.

It can be shown that by ramping the cepstrum and applying linear prediction, that the resulting synthetic spectrum matches not only resonant frequency peaks but also frequency valleys between peaks as well as valleys due to vocal tract zeros such as found in nasals. These spectral matching properties are currently being investigated with the aim of incorporating this method into a complete high quality, low bit rate analysis-synthesis system.

The theoretical details of spectral matching by linear prediction are given in the University of Utah Technical Memo #7500, May 6, 1975. The theory supporting the technique of applying linear prediction to the ramp modulated cepstrum was published in the "Proceedings of the IEEE 1975 Region Six Conference on Communications" by A. Atashroo.

Section 7

Suppression of Noise from an Audio Signal
Tracy Petersen

The problem addressed here is the suppression of additive broad-band stationary noise from a signal whose spectral characteristics vary with time. The solution to this problem has important applications wherever a time varying signal has been polluted with additive stationary noise--possible examples being the interpretation of distress messages or signals intercepted from a noisy environment.

Because the signal of interest varies with time, it is desireable to design a noise suppression filter that continuously adapts itself to the characteristics of the signal. A major difficulty in the realization of such a filter is that its frequency response must change smoothly and in a well behaved manner between successive discrete time estimates of the optimum noise suppression filter.

An important result of this research has been the development of a practical method which facilitates changing the frequency response of a filter smoothly in time, where spectral characteristics are periodically specified in the frequency domain. This filtering scheme represents a new approach to the dynamic filtering of signals, as well as a new application of a filter traditionally used for speech synthesis.

Linear prediction theory has demonstrated the effectiveness of the lattice form digital filter for speech synthesis from a set of speech analysis parameters [1,2]. An implementation of this same lattice filter has been developed, which makes possible the arbitrary and dynamic filtering of the signals, where the frequency response of the filter is modeled as an all-pole transfer function. The implementatoin of this method is as follows.

In the linear prediction analysis-synthesis of speech, the central parameters to the lattice filter, known in the literature as k-parameters [1], are derived from the short time autocorrelation of the speech signal. The frequency response of the lattice filter in this case models the short time spectral envelope of the speech. If a set of k-parameters are derived from the autocorrelation of the impulse response of a filter speicified arbitrarily in the frequency domain as a prototype, the frequency response of the lattice filter controlled by this set of k-parameters will model the prototype filter [3]. Thus, if it is desired to change the spectral characteristics of the filter smoothly in time from an initial to a target configuration, this can be realized by first deriving two sets of k-parameters corresponding to the initial and target filter configurations, respectively, and then smoothly interpolating the k-parameters in time from the initial to the target set. This means the lattice filter will have a new set of control parameters at each sample point in time.

A dynamic version of the Wiener filter has been realized with

this technique, and has been used with success to suppress the surface noise from an old acoustic recording of singing voice. This requires continuously updating the target filter estimate over short time intervals. If filter update intervals are too far apart in time, some unwanted noise will be admitted into the filtered signal. This research has found that noise suppression from a singing voice requires an update interval on the order of 150 milliseconds. Each time the power spectrum of the filter is determined as

$$\frac{\Phi(s+n)-\Phi(n)}{\Phi(s+n)} = \Phi_w$$

where $\Phi(s+n)$ is the power spectrum of the noisy input signal from the particular time frame being analyzed, and $\Phi(n)$ is the estimated power spectrum of the noise (assumed constant). The current Φ_w is mapped by an inverse Fourier transform into the time domain function R_w which is the autocorrelation of the impulse response of $(\Phi_w)^{\frac{1}{2}}$. Robinson's method [4] is used to map R_w into a new target set of K-parameters which determine the filtering characteristics of the lattice filter. A cosine curve has been chosen to interpret between initial and target K-parameters. This avoids instantaneous change in the velocity of the K-parameters causing the frequency response of the lattice filter to converge smoothly to the current target estimate. The noisy input signal is supplied as a driving function to the lattice filter which then filters the input signal with smoothly changing estimates of $(\Phi_w)^{\frac{1}{2}}$.

Present limitations on this dynamic filtering technique reside in the fact that the transfer function is approximated with an all-pole model. Since zero information is not used explicitly in the model, a relatively large number of poles are required should narrow frequency band attentuation be desired. Current work with this filtering scheme has been done with 90 poles and a maximum attenuation level of -24db. It is believed the inclusion of zero information into the approximating transfer function will greatly increase the filtering efficiency of this model.

A means for extracting this zero information is a subject of continuing research within the Sensory Information Processing Group at Utah.

References for Section 7

[1] Morgel, J.D.; Gray, A.H. and H. Wahita. "Linear Prediction of Speech--Theory and Practice." SCRL Monograph No. 10, Sept. 1973.

[2] Itakura, F. and S. Saito. "On the Optimum Quantization of Feature Parameters in the PARCOR Speech Synthesizer." Paper 14, Musashino Electrical Communication Laboratory, N.T.T. Musashino, Tokyo, Japan.

[3] Makhoul, J. "Linear Prediction: A Tutorial Review." Proceedings of the IEEE, Vol. 63, No. 4, April 1975.

[4] Robinson, E.A. "Multichannel Time Series Analysis with Digital Computer Programs." New York:Holden Day (1967).

Section 8

Demonstration: An Isolated-Word Recognition System for Flight Management

Mike Coker

8.1 Introduction

There is currently a program of research at Ames Reeearch Center (NASA) on pilot procedures and pilot-system interfaces [1]. Briefly, the goal is to move toward automation of flight management and air traffic control processes, and simplification of tasks currently performed by flight management personnel. The proposed system would include an onboard computer to perform low-level status appraisement, monitoring, and executive functions. As a medium to input commands and data to the system, speech has been suggested as being most natural and efficient [1].

Toward the goal of designing a system for isolated word recognition, a modified version of an algorithm presented by Itakura [2] has been implemented. Results of experiments performed ueing thie algorithm indicate that a high recognition accuracy can be obtained.

8.2 Results

The method described below, together with a non-linear time-warping scheme implemented with a dynamic programming algorithm [2], was initially applied to a ten word vocabulary consisting of the digits 0 through 9. The single speaker accuracy achieved wae 100%. During this experiment, it was discovered that a high rate of

recognition accuracy can be obtained on these words using far less signal information than would be required to reconstruct the speech with high intelligibility, as required in a speech analysis-synthesis system.

This work-recognition scheme was subsequently applied to a vocabulary of 107 words, consisting of the original ten digits plus 97 flight commands. The single speaker accuracy achieved varied from 95 to 98% depending on the number of training utterances for each word. Computation time was about twelve seconds per word.

Two other vocabularies were also tried. In the case of the ten digits plus the alphabet, 89 to 95% accuracy was obtained for a single speaker. In the case of twenty-five word pairs from the Diagnostic Rhyme Test, a standard vocoder intelligibility test, accuracies ranged from 78 to 88%.

8.3 Analysis

The most important considerations in a word-recognition system are: (1) recognition accuracy; (2) computation time; and (3) reference pattern storage requirements.

Given that an effective distance measure exists for comparing short-time power spectra, it is apparent that recognition accuracy depends heavily on the reliability of the algorithms chosen for detection of the beginning and ending of each utterance (endpoint detection) and for proper alignment of the speech sounds between two utterances (time- warping). Endpoint detection is normally done by

comparing the energy (zeroth autocorrelation coefficient) of euccessive speech segments to some threshold value. Since fricated phonemes often form word endpoints, which are sometimes the only difference between two different words (i.e., singulars vs. plurals), endpoint detection is done on the differencee epeech signal. Because power spectra derived from fricated phonemes normally have much less total energy than voiced sounds (while having relatively more energy at high frequencies), the differencing, which emphasizes high frequencies and attentuates low frquencies [4], allows more accurate endpoint detection on both voiced and unvoiced sounds.

The non-linear time-warping algorithm matches epeech sounds between words to be used so that variations in the way they are spoken does not affect recognition accuracy. It is apparent that the eame distance measured which is used to compare short-time power epectra between words, could also be used to separate speech sounds within a single word. When this is done, the reference pattern can be constructed using only significantly different power spectra, thereby reducing the pattern size and thus the amount of computer etorage and computation necessary for pattern matching. Also, sound eeparation within a word can be done using a lower order linear prediction model than the actual reference pattern formation, so that considerable computation time can be saved.

Another desirable feature of the word-recognition system is the capability to form a reference pattern based on more than one

utterance of each vocabulary word. This enables the system to follow long-time variations in word pronunciation.

The system was implemented in real time on the Univereity of Utah's single-user PDP-10 computer.

8.4 Method

The linear prediction method of signal analysis provides a convenient distance measure between short-time (20-30 msec) all-pole power epectra [2,3], thereby permitting segment-wise comparisons between isolated words. Itakura [2] has used this distance measure to attain 97.3% recognition accuracy on a vocabulary of 200 words.

The model for linear prediction analysis assumes that each signal sample in a segment of N samples can be closely predicted by a linear combination of the preceding p samples for $p < N/2$, thus:

$$s_n = \sum_{k=1}^{p} a_k s_{n-k} + e_n \qquad n=0,1,2,\ldots N-1$$

where $[a_k]$ $k=1,2,\ldots p$ are known as the predictor coefficients and $[e_n]$ $n=1,2,\ldots N-1$ is the prediction error at each point. The predictor coefficients are found by minmizing the linear prediction residual (LPR), defined by

$$LPR = \sum_{n=0}^{N-1} e_n^2 .$$

For a given segment of speech, the prediction error sequence is given by

$$e_n = s_n - \sum_{k=1}^{p} a_k s_{n-k} .$$

For the same set of predictor coefficients $[a_k]$ and a difference set of speech samples $[\hat{s}_n]$ n=1,2,... N-1, the associated prediction error sequence is given by

$$\hat{e}_n = \hat{s}_n - \sum_{k=1}^{p} a_k \hat{s}_{n-k}$$

Therefore, a measure of similarity between the two speech segments represented by $[s_n]$ and $[\hat{s}_n]$ is given by the "closeness" of the two numbers

$$LPR_1 = \sum_{n=0}^{N-1} e_n^2 \quad \text{and} \quad LPR_2 = \sum_{n=0}^{N-1} \hat{e}_n^2 ,$$

or since $LPR_1 \leq LPR_2$, how close the ratio

$$\frac{LPR_2}{LPR_1}$$

is to unity.

It can be shown [3] that this distance measure represents a comparison between the power spectrum $|A(\omega)|^2$ of the inverse filter defined by the predictor coefficients $[a_k]$ (extracted from the first speech segment) and the power spectrum $P(\omega)$ of the second speech segment $[\hat{s}_n]$, in this way:

$$\sum_{n=0}^{N-1} \hat{e}_n^2 = T/2\pi \int_{-\pi/T}^{\pi/T} P(\omega) |A(\omega)|^2 d\omega$$

where T is the sample interval.

References for Section 8

[1] Wempe, T.E. "Flight Management--Pilot Procedures and System Interfaces," Technical Report, Ames Research Center, NASA.

[2] Itakura, F. "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Transactions on Acoustics, Speech and Signal Processing, February 1975.

[3] Boll, S.F. "Waveform Comparison Using the Linear Prediction Residual," University of Utah Computer Science Technical Memorandum, TM #7500, May 1975.

[4] Makhoul, J.I. and Wolf, J.J. "Linear Prediction and The Spectral Analysis of Speech," BBN Report No. 2304, August 1972.

Section 9

Word Recognition in Continuous Speech
Richard W. Christiansen

9.1 Introduction and Background

There are rather extensive efforts already underway in the areas of automatic speech recognition, speech understanding, and word recognition using digital waveform processing [1]. Even a casual reading of this recent overview of the current work will convince one of the magnitude and extreme difficulty of the general speech recognition problem. Some of the key problems associated with this work are:

1. Segmentation of incoming speech into:
   a. Phonemes
   b. Syllables
   c. Words
2. Defining beginnings and endings of words
3. Use of syntax and semantics
4. Classification of speech segments
5. Time registration--usually nonlinear

However, the work described in this section is fundamentally different, and is directed toward achieving a different objective than the general speech recognition work. This difference allows one to choose other ways of solving the problem which eliminate most of the difficulties inherent in the general solution. Of the five problems referred to, only the time registration problem remains.

Here we are not trying to identify an input work or determine its meaning, but merely determine if a reference word ever occurs using simple waveform matching techniques. This technique may be described as a time flexible template matching approach in which the input speech is analyzed one frame at a time, and time registered to the stored reference template using a dynamic programming algorithm. At the completion of the analysis using each frame, a test is made to determine if that frame formed the last frame of the word corresponding to the template word. A frame similarity function, a threshold, a parameter for weighting previous data, and a time distortion penalty constant are used in the dynamic programming algorithm which is described later. Although the final result of this effort may not yield a fully automatic system, it should still produce a practical working system which uses some human interaction but still accomplishes the main objective.

9.2 Objective

There were three main objectives of this research. They were as follows:

1. Given a recording of someone talking, i.e., a continuous speech recording, and a reference example of the same person speaking a word from the continuous speech text, then identify whenever the reference word occurs in the recording using digital signal processing techniques.

2. Provide a statistical evaluation of the

performance of the system, an explanation of why it works when it does and fails when it does which can be related back to the acoustic model for speech, a measure of the robustness of the system which characterizes the sensitivity to "noise," small changes in parameters, "quality" of recording system and environment, etc., and in general, advance the understanding of speech waveform matching using LPC parameters.

3. Provide a preliminary evaluation of the system for the case of using multiple templates both from the same speaker and from several different speakers.

9.3 Basic Word Finding Algorithm

As a possible basic approach to locating a word in continuous text, one could choose to segment the incoming speech into a string of phonemes, classify the phonemes, and then identify the word based on this classification. This approach has several major problems, among which the most serious are segmentation and classification of phonemes.

As another approach, one could view a word as a waveform with certain fundamental structure or spectral shape, and then design a filter matched to that wave shape. This approach, common in radar work, is called the matched filter approach. The problem with this approach is that of time registration. That is, any word may be spoken at a different speed each time it is spoken and, in fact,

with different parts of the word spoken at different speeds each time.

To use this approach one needs a measure of spectral matching and some type of nonlinear time warping procedure to somehow measure how well the template word matches the word in the continuous speech.

In searching for clues to solving these problems, it was discovered that Bridle [2] had achieved 85 percent success at word recognition using a 19-channel vocoder and a rather clever dynamic programming scheme. His procedure uses a frame similarity function derived by performing a cosine transformation on a 19-point logarithmic, short-term power spectrum. This yields some spectrum shape coefficients which are then weighted using empirically derived weighting coefficients. A squared distance between frame a and b is then formed and transformed into the frame similarity number c so that $0 \leq c \leq 1$. He then performs the time registration using dynamic programming. To do this he defines a local similarity function AR which gives a measure of how similar the incoming speech is to the template on a frame-by-frame basis. The local similarity function is a weighted sum of the frame similarity functions along the registration path to the current position.

This local similarity function includes a factor which applies a penalty if there is time distortion and no penalty if there is not. (See Apppendix A.)

There were a number of attractive points in Bridle's work, some of which are listed below.

1. The dynamic programming routine is simple.

2. Any other frame similarity function which is bounded by 0 and +1 could be used.

3. It performed the nonlinear time registration with one input frame at a time.

4. It worked quite well using the channel vocoder.

It was therefore decided to use the inverse of the linear prediction ratio given in Equation 58 in a memo by Boll [3] as the frame similarity function and to apply the dynamic programming exactly as given by Bridle.

Now the following things should be noted concerning this algorithm:

1. It is a recursive algorithm which computes Imax values for an array AR, where Imax = number of analysis frames in the template.

2. SIM is a function which describes the similarity between a frame of the template and the incoming frame.

3. AR is a local similarity function which is really a weighted sum of the frame similarity functions along the path to the current position.

4. The local similarity at the first point on a path is defined to be equal to the frame similarity at that point.

5. A time distortion penalty is applied so as to allow poorly matching frames, providing they occur between well matched regions, and to provide a control parameter to limit or reduce the contributions due to "excessively large" time distortions.

6. At the beginning of the process, the AR array is zero. Then if the input speech begins to match the template, nonzero values begin to propagate through the AR array. If AR(Imax, j) $\neq 0$, then AR(Imax, j) is identified as the ending frame of the word. If instead the input speech begins to match the template and then does not match well for succeeding frames, the recursive nature of the algorithm successively sets the elements of the AR array back to zero.

7. The array AR contains all necessary information about previous input speech and how well it has matched.

9.4 Current Status of This Research

A preliminary version of this algorithm has been implemented on the PDP-10 system and limited testing has been conducted using words spoken in isolation as well as continuous speech. This implementation is shown in block diagram form in Figures 9.1 and

9.2, with the algorithm and basic Fortran program given in Appendix A. The program runs in about 12 times real time.

For preliminary evaluation purposes, ten utterances of each of the two words "octopus" and "carnival" and six utterances of the word "interrupt" were digitized and recorded in isolation. In addition, approximately two minutes of text from a PDP-11 manual was digitized and recorded. A single occurrence of each of the words "program," "processor," and "priority" was located in the text to be used as templates to locate other occurrences of the words.

Frequency of occurrence of these words in the text is given in Table 1.

Table 1. Word list.

Word	Occurrence
Interrupt	18
Program	10
Processor	7
Priority	6

The results are summarized in Table 2, where all template words are spoken by the same speaker as the "text."

Here it should be noted that the length of the interrupt template is approximately .6 seconds while the continuous speech versions were as short as .3 seconds.

In contrast, Bridle, using templates taken from the text reports 85 percent hits with six false alarms per hour. To achieve this, he used as many as eight different thresholds for each analysis and had five different speakers. At this point, there is no realistic basis for comparison since I have no reasonable or even comparble statistics.

In addition to the above results, a test run was also made on a recording by a second speaker using the template from the first speaker and the same Q, G, and RK values. The results were 18 hits, 0 misses, and 2 false alarms.

Table 2. Preliminary experimental results.

Word	Hits	Misses	False Alarms
	From Text		
Program	10	0	0
Priority	6	0	0
Processor	7	0	0
Interrupt	18	0	0
	Q = .275		
	G = .6		
	k = .5		
	Isolated Words		
Carnival	10	0	0
Octopus	10	0	0
Interrupt	6	0	0

All these prelinary results clearly demonstrate the soundness and feasibility of this basic approach. The preliminary success using a template from a different speaker also suggests great promise for solving the problem using templates from the same speaker or from several different speakers. In other words, it may be possible to construct a set of templates of a word spoken several different ways by several different speakers which would allow identification of the word using text from many other speakers. This procedure will be detailed in the next report, along with the results to date.

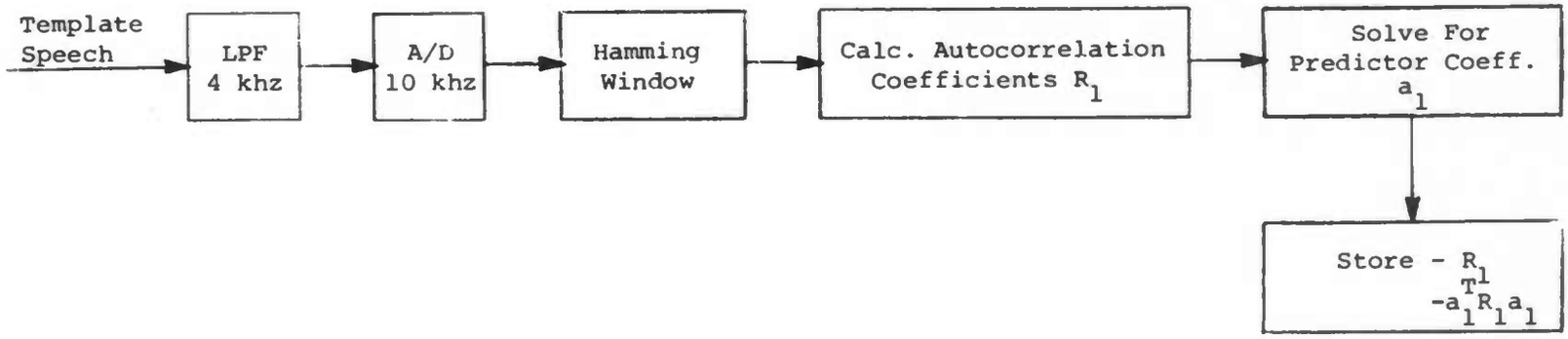


Figure 9.1 -- Analysis for Reference Template

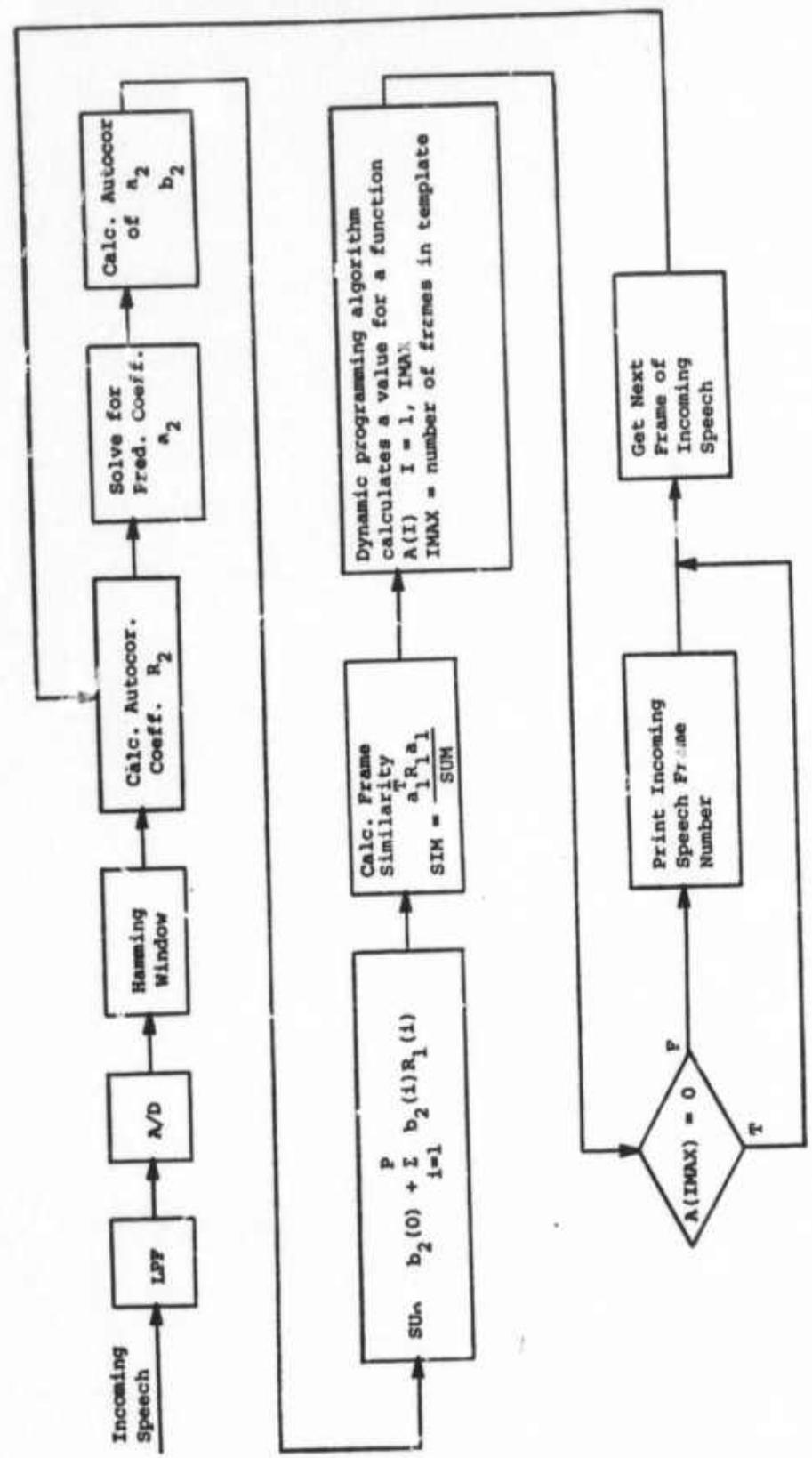


Figure 9.2 -- Analysis of Incoming Speech

References for Section 9

[1] IEEE Transactions on Acoustics, Speech, and Signal Proceeeing, February 1975.

[2] Bridle, J.S. "An Experimental Automatic Word Recognition System." Interim Report, Joint Speech Reeearch Unit, Middleeex, England.

[4] Private discussions with M. Coker and S.F. Boll.

[5] Duda, R.O. and P.E. Hart. "Pattern Classification and Scene Analyeis." New York:John Wiley and Sons (1972).

Appendix A

Word Finding Algorithm

Let a frame similarity function, SIM, be defined by:

$$SIM = \frac{a_1^T R_1 a_1}{a_2^T R_1 a_2}$$

where

R_1 = autocorrelation coefficient matrix for a frame of windowed speech samples for the template

a_1 = linear predictor coefficient vector for a frame of the template speech

a_2 = linear predictor coefficient vector for the incoming speech frame

Then for the dynamic programming algorithm, we can define a function AR(i,j) as follows:

$$AR(i,j) = \max_{(a,b)} Step[(AR(i-a, j-b), SIM(i,j), K(a,b)]$$

AR(I,j) = max [if SIM(i,j) < C, then 0, else SIM(imj),
Step (AR(I,j - 1), SIM(I,j), RK)]

where

Step (AR, SIM, RK) = If AR = 0, then 0

else if (1 - G)AR + GRKSIM < Q, then 0

else (1 - G)AR + GRKSIM

(a,b) takes values (1,0), (1,1), (0,1)

K(a,b) takes corresponding values RK, 1, RK

$0 \leq Q \leq 1$ Q = threshold for similarity

$0 \leq G \leq 1$ G = previous data weighting "time const"

$0 \leq RK \leq 1$ RK = time distortion penalty factor

Section 10

Speech Processing to Remove Noise and Improve Intelligibility

Michael Callahan

A new method of speech processing is being investigated which has applications to noise removal and intelligibility improvement. The method is based on homomorphic processes which have previously been used for picture deblurring and deresonation of acoustic signals. The basic method can be summarized as follows:

1. The short-time spectrum of the original signal is calculated. The short-time spectrum is two-dimensional and shows the frequency content of the signal as a function of time.

2. The short-time spectrum is altered to remove effects due to noise or to amplify speech features using methods such as thresholding or two-dimensional linear filtering.

3. A new acoustic signal is reconstructed from the modified short-time spectrum.

This type of processing is often more effective than conventional methods because of the inherent flexibility of two-dimensional processing, and because of the similarity of the method to processing performed by the human auditory system.

Short-time spectrum processing has been quite successful in two

initial experiments: removal of background noise (signal to noise ratio about 30db), and removal of high level signals with strong harmonic structure, such as 60Hz square wave noise (signal to noise ration about -26db). Both of these experiments, which were discussed in some detail in the last report, are complete.

Present research concerns more general applications of this process. The first step is to develop the capability to extract from the short-time spectrum the acoustic features of speech known to be important to perception. This ability is important for several reasons: a) it will give a better understanding of the effect of noise removal processes on the underlying speech; b) it may provide the basis for pre-processing speech to enhance important features so that the speech will be more intelligible in a noisy environment; and c) it may suggest improved compression-expansion techniques when the speech must be passed through a noisy channel.

An example of this process is shown in Figures 10.1, 10.2 and 10.3. Figure 10.1 is the short-time spectrum of the words "we were away a year ago." The vertical axis is frequency (0-5000 Hz), the horizontal axis is time (1.6 sec), and brightness is proportional to the magnitude of the short-time spectrum. Figure 10.2 shows the speech formants for the words in Figure 10.1. Formants are prominant peaks in the speech spectrum caused by resonances in the vocal tract for the speaker--the frequency, inteneity and bandwidth are important to intelligibility and naturalness. The formants were obtained from the sentence of Figure 10.1 by filtering the logarithm

of the magnitude of the short-time spectrum. Figure 10.3 shows the original sentence after the formants have been enhanced. Preliminary experiments indicate that speech processed in this manner is more intelligible in a noisy environment.

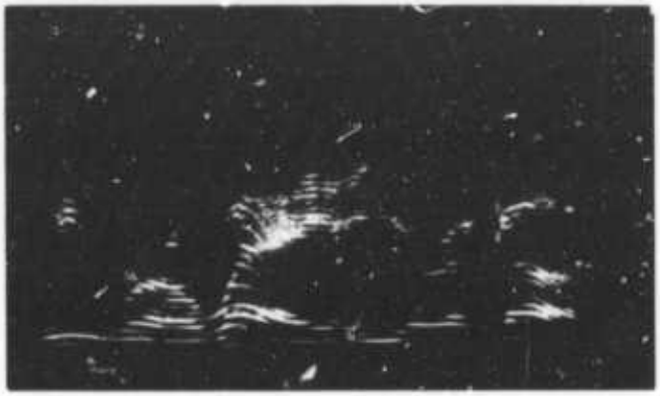


Figure 10.1 -- Short-time spectrum of original sentence.

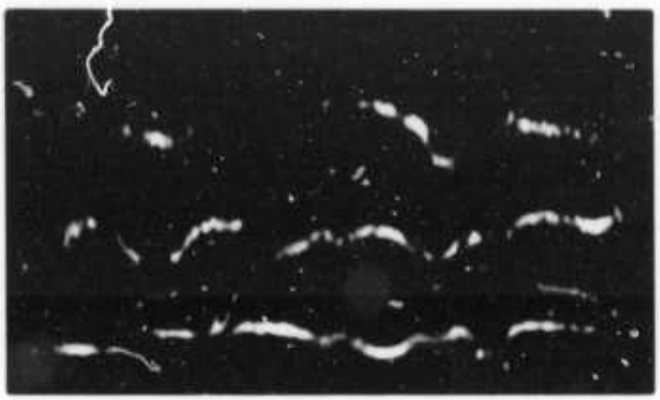


Figure 10.2 -- Formants extracted by two-dimensional filtering of log of short-time spectrum.

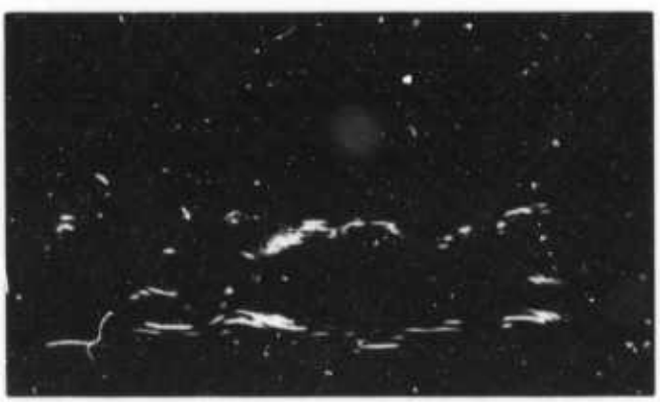


Figure 10.3 -- Short-time spectrum of new sentence with enhanced formants.

Section 11

Linear Predictive Coding with Zeros and Glottal Wave

L.K. Timothy

11.1 Introduction

The quality of synthesized speech in Linear Predictive Coding (LPC) as presented by Atal [1], suffers for at least two reasons:

1. Improper treatment of the glottal pulse or wave.

2. Lack of zeros in the all-pole mathematical model which are needed for nasal sounds and radiation effects.

For comparative purposes, Figure 11.1 presents a natural voiced speech waveform, while Figure 11.2 presents a corresponding waveform synthesized by the Atal method. Figure 11.3 presents a corresponding waveform synthesized by the proposed method of this paper.

For speech synthesis, rather than exciting an all-pole filter with an impulse as Atal did, the filter is excited with an estimated glottal wave sequence modified by zeros $\{E_i\}$ as indicated in Equation (1).

$$s_n = \sum_{i=1}^{p} a_i \, s_{n-i} + E_n \tag{1}$$

$$E_n = \sum_{i=1}^{r} b_i \, g_{n-i} \tag{2}$$

The b_i coefficients are the zero coefficients, and the sequence $\{g_i\}$ is an assumed glottal wave. The sequence $\{E_i\}$ is actually the error signal; however, in this application only that part of the error signal when the glottis is judged to be open is used, the remainder set to zero. Instead of transmitting the error signal, the b_i coefficients (or coefficients for the zero polynomial) are estimated based upon an assumed glottal wave sequence $\{g_i\}$ and transmitted. The error signal is recreated in the receiver from the b 's and the assumed $\{g_i\}$.

The sequence of mathematical operations for voiced speech follows. Unvoiced speech is treated as recommended by Atal:

1. The interval of time during a pitch period when the glottis is judged to be open is estimated which requires pitch synchronous information. Figure 11.1 illustrates the idea.

2. A weighted least squares estimate of the predictor (or reflection) coefficients is made based upon only those portions of the speech wave when the glottis is closed.

3. An error signal is formed as indicated in Equation (1) using actual speech wave data $\{s_i\}$.

4. The zero coefficients b_i are estimated from only that part of the error signal that corresponds to when the glottis is judged to be open as illustrated in Figure 11.4.

Although zero coefficients and the error signal must be calculated in addition to the usual Atal calculations, fewer predictor coefficients need be calculated. Consequently no significant increase in the number of computer operations is expected, and real time implementation is expected. Real time synchronous pitch detection, which operates with few errors, was reported in Reference 2.

11.2 Zeros and the Glottal Wave

In the mathematical model of the speech wave presented in Equations (1) and (2), the zeros act only on the forcing function, the glottal wave. Consequently, according to the model, the zeros play an active role only when the glottis is open. The zeros only indirectly affect the ballistic or free portion of the speech wave when the glottis is closed by modifying the glottal wave which in turn adjusts the initial condition of the ballistic portion at the point when the glottis closes. The error signal, therefore, is approximately the glottal wave modified by zeros of the system.

11.3 Glottal Interval

The interval of time during a pitch period when the glottis is judged to be open terminates at the absolute maximum value of the speech wave as indicated in Figure 11.1. The beginning of the glottal interval is judged to be two zero crossings prior to the terminal point. Flanagan [3], estimates the interval to vary from 30 to 70 percent of the pitch period which will usually be different

from this assumption. However, the quality of the synthesized speech, based upon the above assumptions, is very good.

11.4 Suppression of Distorted Data

In the frequency domain analysis of speech, the suppression of unwanted data is a very difficult problem. By utilizing weighted least squares estimation in a time domain analysis, the suppression of unwanted data becomes a straightforward process as will be demonstrated below.

In LPC, a linear regression analysis using least squares estimation is used to estimate predictor coefficients in vector form.

$$\underset{\sim}{a} = \begin{bmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_p \end{bmatrix} \tag{3}$$

from sampled speech,

$$\underset{\sim}{s}_n = \begin{bmatrix} s_n \\ s_{n-1} \\ . \\ . \\ . \\ s_{n-q} \end{bmatrix} \tag{4}$$

The least squares estimate of the predictor coefficients can be written as

$$\hat{\underset{\sim}{a}} = \left(\underset{\sim}{H}^T \underset{\sim}{H}\right)^{-1} \underset{\sim}{H}^T \underset{\sim}{s}_n \tag{5}$$

where the $\underset{\sim}{H}$ matrix contains sampled data as follows:

$$\underset{\sim}{H} = \left[\underset{\sim}{s}_{n-1} \;\vdots\; \underset{\sim}{s}_{n-2} \;\vdots\; \cdots \;\vdots\; \underset{\sim}{s}_{n-p}\right] \tag{6}$$

The $\underset{\sim}{H}^T\underset{\sim}{H}$ matrix is the autocorrelation of autocovariance matrix.

The weighted least squares estimate of the predictor coefficients can be written as

$$\hat{\underset{\sim}{a}} = \left(\underset{\sim}{H}^T\underset{\sim}{R}^{-1}\underset{\sim}{H}\right)^{-1} \underset{\sim}{H}^T\underset{\sim}{R}^{-1}\,\underset{\sim}{s}_n \tag{7}$$

where $\underset{\sim}{R}$ normally would be the covariance matrix on the estimation error.

$$\underset{\sim}{R} = E\left|\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}^T\right|$$

$$\underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_q \end{bmatrix} \quad ; \; \varepsilon_i = s_n - \sum_{i=1}^{p} a_i\, s_{n-i}$$

If $\underset{\sim}{R}^{-1} = \underset{\sim}{I}$, the identity matrix, then the weighted estimate, Equation (7), would be identical to Equation (5), the unweighted least squares estimate. If one chooses to throw out some data and retain other data with equal weighting, the $\underset{\sim}{R}^{-1}$ can be chosen as follows:

$$\underset{\sim}{R}^{-1} = \left[\begin{array}{c|c|c} \underset{\sim}{I}_1 & 0 & 0 \\ \hline 0 & 0 & 0 \\ \hline 0 & 0 & \underset{\sim}{I}_3 \text{ etc.} \end{array}\right] \tag{8}$$

The $\underset{\sim}{R}^{-1}$ expression in Equation (8) is diagonal with ones or zeros on the diagonal. The ones correspond to the data to be retained, and the zeros correspond to the data to be discounted.

If the $\underset{\sim}{H}$ matrix is partitioned to correspond to Equation (8) as follows:

$$\underset{\sim}{H} = \begin{bmatrix} \underset{\sim}{H}_1 \\ \underset{\sim}{H}_2 \\ \vdots \\ \underset{\sim}{H}_r \end{bmatrix}$$

the autocovariance matrix $H^T H$ can be written as

$$\underset{\sim}{H}^T\underset{\sim}{H} = \underset{\sim}{H}_1^T \underset{\sim}{I}_1 \underset{\sim}{H}_1 + \underset{\sim}{H}_2^T 0 \underset{\sim}{H}_2 + \underset{\sim}{H}_3^T \underset{\sim}{I}_3 \underset{\sim}{H}_3 + \ldots$$

$$= \underset{\sim}{H}_1^T \underset{\sim}{H}_1 + \underset{\sim}{H}_3^T \underset{\sim}{H}_3 + \text{etc.} \qquad (9)$$

Equation 9 generally requires fewer multiples than does $\underset{\sim}{H}^T\underset{\sim}{H}$ when all of the data are retained in the usual LPC situation. The $\underset{\sim}{H}^T\underset{\sim}{R}^{-1}\underset{\sim}{e}_n$ terms are calculated as a subset of Equation (9).

11.5 Estimation of Zeros

The polynomial coefficients for the zeros b_i can be deconvolved from the glottal portion of the error wave as represented in Equation (2) as follows. Equation (2) can be expressed as a system

of equations expressed in Equation (10):

$$\begin{bmatrix} E_n \\ E_{n-1} \\ \cdot \\ \cdot \\ \cdot \\ E_{n-q} \end{bmatrix} = \begin{bmatrix} g_{n-1} & g_{n-2} & \cdots & g_{n-r} \\ g_{n-2} & g_{n-3} & \cdots & g_{n-r-1} \\ & & & \\ & & & \\ g_{n-q-1} & g_{n-q-2} & \cdots & g_{n-q-r} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_r \end{bmatrix} \tag{10}$$

A glottal wave is assumed which forms the g_{ij} elements. As Equations (1) and (2) indicate, p poles and r zeros are employed in the model. In more compact matrix form, Eq. (10) can be written as

$$\underset{\sim}{E} = \underset{\sim}{G}\underset{\sim}{b} \tag{11}$$

where $\underset{\sim}{E}$, $\underset{\sim}{G}$, and $\underset{\sim}{b}$ correspond to the matrices in Eq. (10). It is assumed that $q + 1 > r$. Consequently least squares calculation of the zero coefficient may be used which is

$$\hat{\underset{\sim}{b}} = \{\underset{\sim}{G}^T\underset{\sim}{G}\}^{-1}\ \underset{\sim}{G}^T\underset{\sim}{E} \tag{12}$$

Since the glottal wave shape is assumed known, the elements, d_{ij}, of $\underset{\sim}{D} = \{\underset{\sim}{G}^T\underset{\sim}{G}\}^{-1}\ \underset{\sim}{G}^T$ can be precalculated and stored in computer memory such that

$$\hat{b}_i = \sum_{j=1}^{q} d_{ij}\ E_j \tag{13}$$

The glottal wave is assumed to be a raised cosine wave.

References for Section 11

[1] Atal, B.S. and S.L. Hanauer. "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave." Journal of the Acoustical Society of America. Vol. 50 (1971), pp. 637-655.

[2] Miller, N.J. "Pitch Detection by Data Reduction." IEEE Symposium Record on Speech Recognition, April 15-19, 1974, Carnegie-Mellon University, Pittsburg, Pennsylvania, T-9, pp. 122-130.

[3] Flanagan, J.L. "Speech Analysis and Synthesis, and Perception." Springer-Verlag (1972), p. 13.

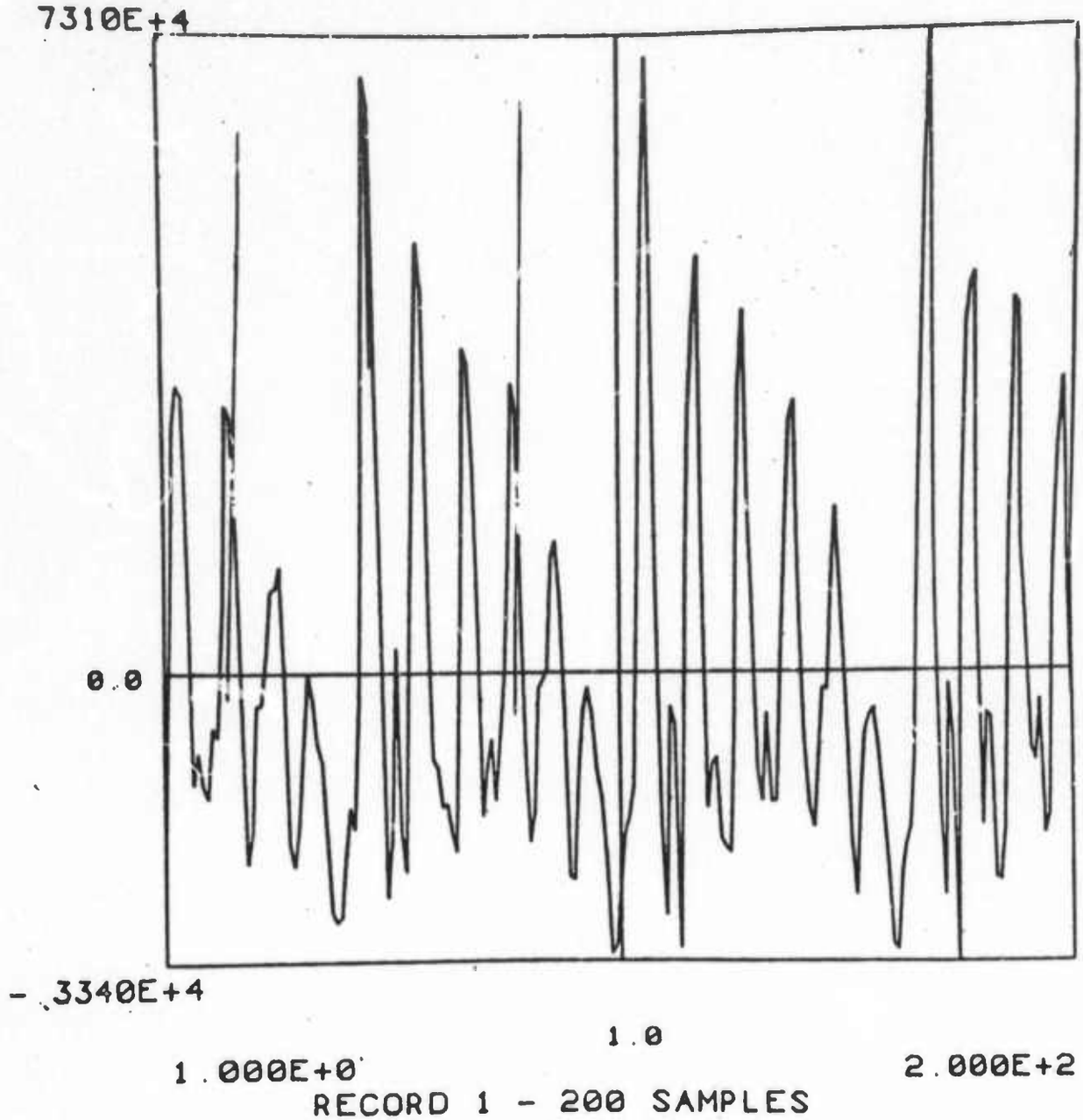


Figure 11.1 -- Original waveform

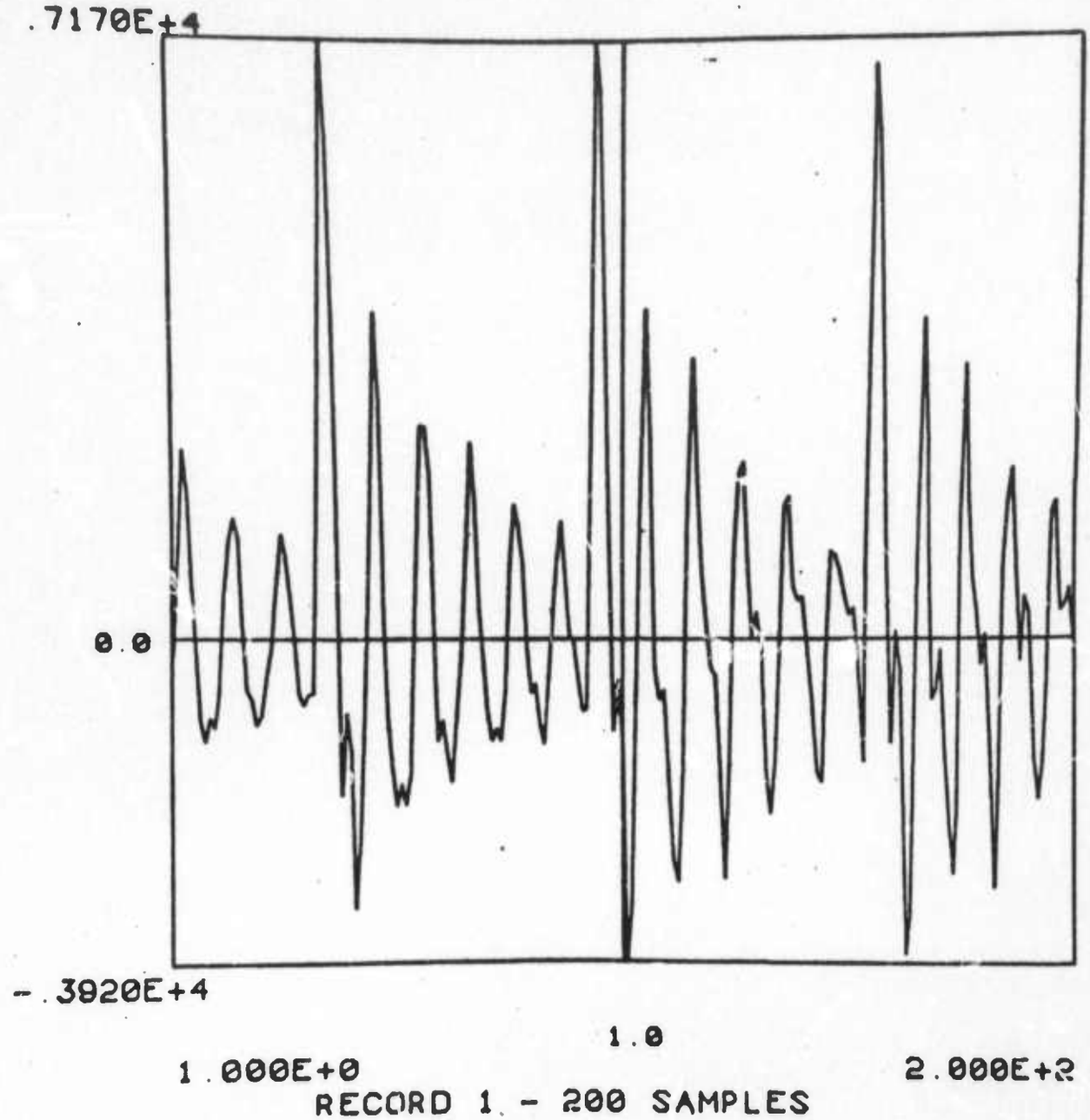


Figure 11.2 -- Synthesized waveform using impulse excitation

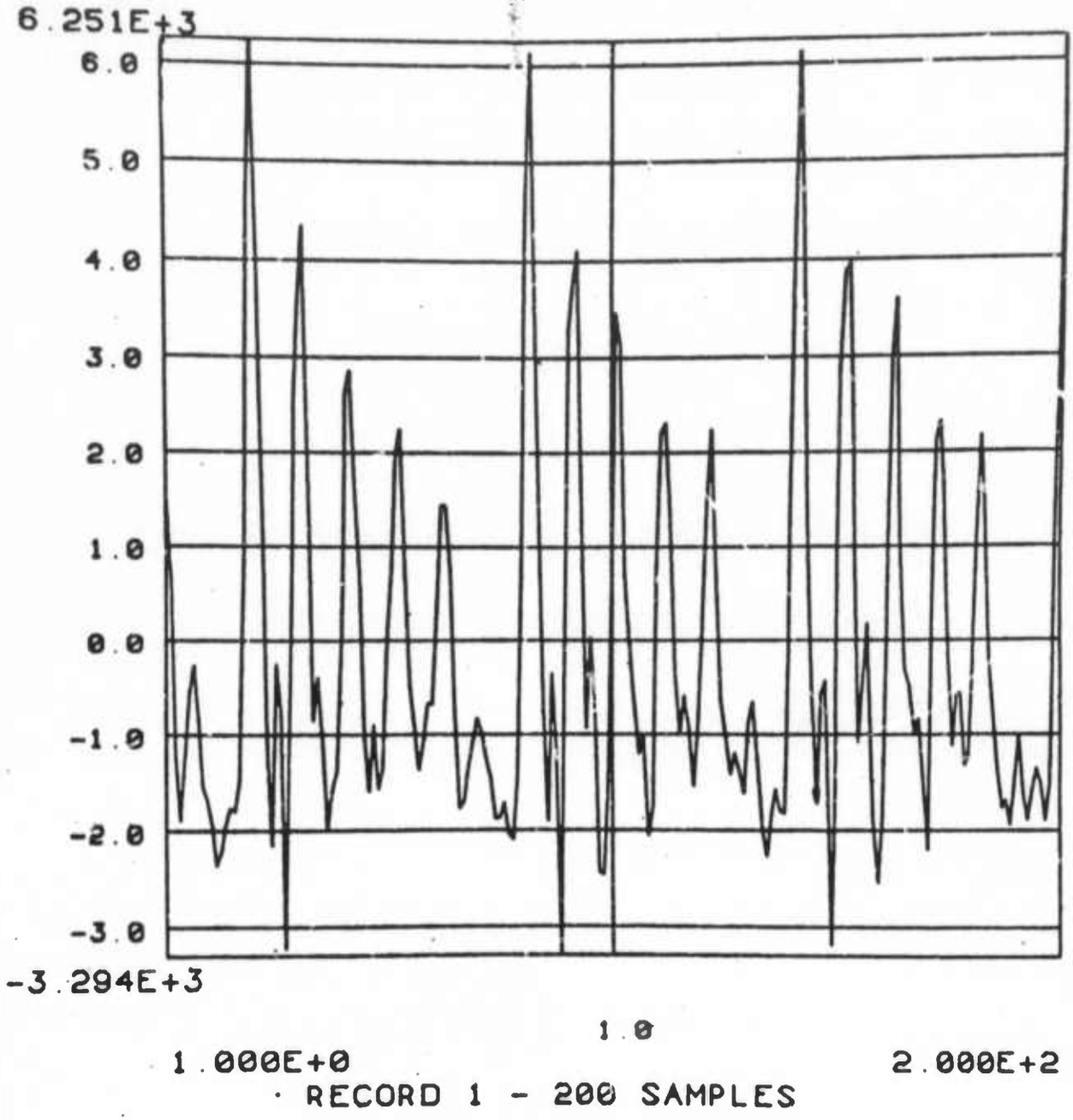


Figure 11.3 -- Synthesized waveform using a partially nulled error signal for excitation.

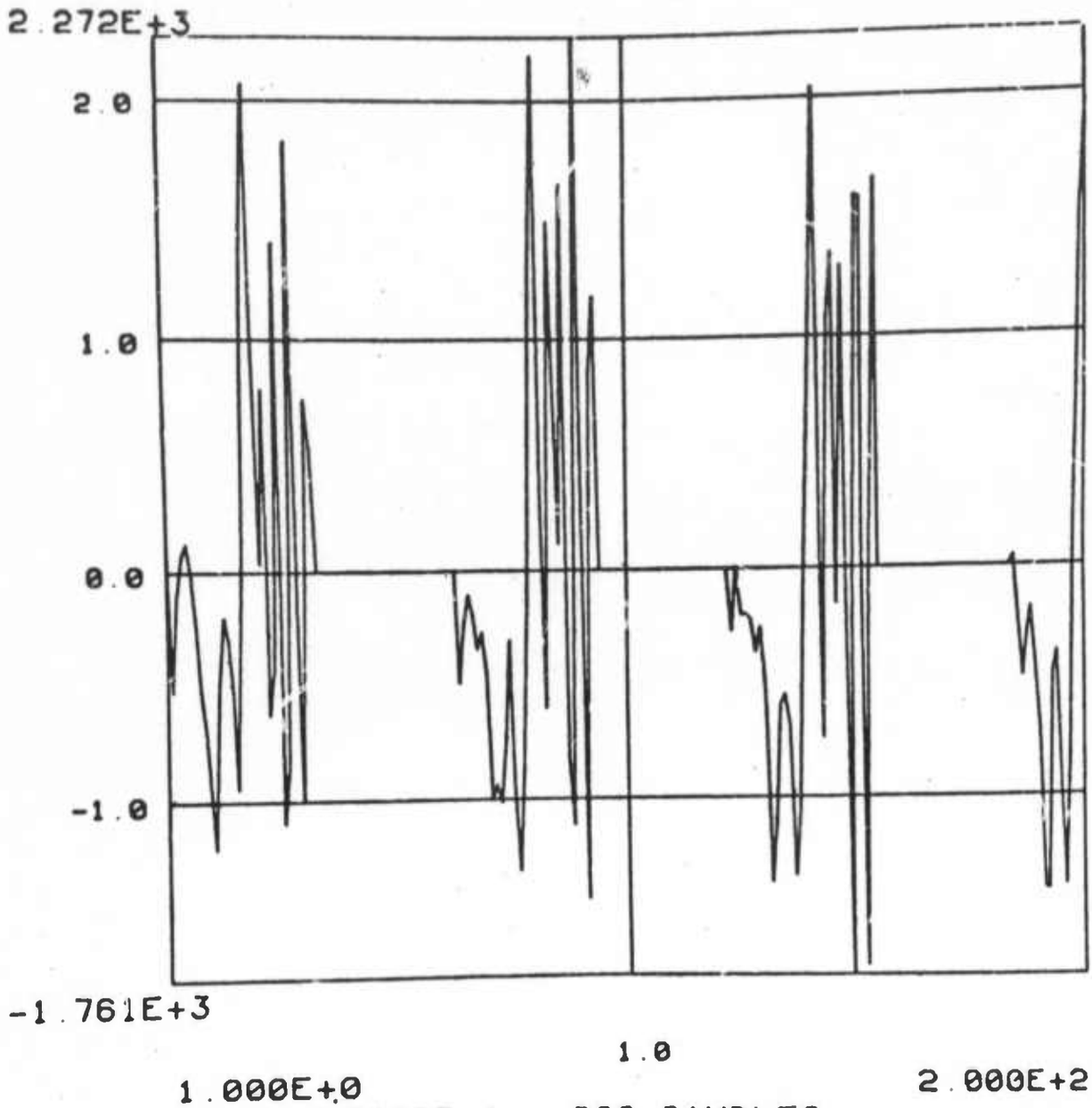


Figure 11.4 -- Partially nulled error signal.

Section 12

Phase Estimate of a Linear System by Blind Deconvolution

Chien H. Chiang

12.1 The Problem Description

In spite of the Ohm's Law on acouetics [1], many studies [2,3] have shown that the phase distortion does cause audible deterioration of speech and music quality. We, therefore, make the hypothesis that the residual reverbrant characteristics in an old recording, after having been deconvolved for magnitude compensation, is caused by the phase distortion associated with the old recording technology. Hence, it is possible to further improve the audio quality in an old recording by proper phase compensation (or deconvolution) if one can also estimate the phase of the recording system which was used in making the recording.

Since the only data available is the phase distored waveform, both the original waveform and the distorting system are unknown, this task of estimating the phase of the linear system can also be regarded as "blind deconvolution."

12.2 The Theoretical Background

The recorded signal v(t) is related to the original waveform e(t) by

$$v(t) = e(t) \circledast h(t) \qquad (1)$$

where h(t) represents the linear system used in making the recording.

Taking the Fourier transform producing

$$V(f) = S(f) \cdot H(f) \tag{2}$$

Further, taking the complex logarithms of both sides of (2), one obtains

$$\log|V(f)| = \log|S(f)| + \log|H(f)| \tag{3}$$

and

$$\angle V(f) = \angle S(f) + \angle H(f) \tag{4}$$

Equation (3) relates the magnitudes and (4) represents the relation between the phases. The main objective here is to estimate $\angle H(f)$.

One approach might be to take several recordings made with the same recording equipment, calculate the phase, put each one in the form of (4) and average both sides of these equations across all of the recordings, namely

$$\frac{1}{N}\sum^{N} \angle V_i(f) = \frac{1}{N}\sum^{N} \angle S_i(f) + \angle H(f) \tag{5}$$

where N is the number of the recordings. If there were enough recordings and the speech or signing on each recording were quite different, then one might expect, according to the central limit theorem, that the right hand side of (5) will converge to $\angle H(f)$. The difficulty is that it is hard to come by enough recordings which are known to be made with the same equipment. The way to get around the problem is to chop up one recording into sections and take the phases of these sections as the ensemble on which to carry out the averaging. Thus, it might be possible to estimate the phase of the linear system by a three step process: 1) average the phase of the distorted waveform; 2) average the phase of a similar acoustic wave which has not been phase distorted (probably by using a good recording of "identical" speech or singing); and then 3) subtracting between the two.

12.3 The Difficulties Encountered

Two difficulties arise when one tries to implement the process described above:

a) The Phase Unwrapping.--The digital computer utilizing the conventional four quadrant inverse tangent function routine computes only the principle value of the phase. Unfortunately, the sum of two principle values does not give the principle value of the sum. So one has to, somehow, find the actual phases before he averages them. An algorithm by Schafer [4] is to decide the number of jumps of 2π so that one can obtain the actual phase from its principle value version. The process is called phase unwrapping and both

Schafer and Oppenheim [5] have discussed this topic elsewhere. The phase unwrapping scheme of Schafer will not work when there are sudden energy dips in the spectrum. Unfortunately, this is just the case when the time waveform is periodic, as is true for speech. The dips cause the spectrum to become "discontinuous" in some frequencies.

To get the best unwrapping result using Schafer's method, the phase has to be computed from a data frame so small that the fine details (the energy dips) will not appear in the spectrum.

b) The Reference Points in the Time Waveform.--Many people have noticed that a slight shift in the relative position of the window with respect to the waveform causes a tremendous change in the unwrapped phase [6]. This phenomenon is undesirable especially when the statistic property of the phase is of major concern.

To circumvent the effect of this change, we "synchronize" the waveform prior to the taking of FFT in such a manner that the waveform will always have the peaks located at the beginning point of the FFT array. In other words, we take FFT in a synchronous fashion with respect to the peak of the time waveform.

It turns out that using this synchronization scheme, the phase calculated becomes statistically stabilized. Thus, we hope to avoid the bad statistic behavior noted by other researchers.

12.4 The Experiment and The Results

An initial experiment was carried out to illustrate the method of estimating the phase of the linear system given only the distorted waveform.

A speech passage by the male subject number 1 reading from a text is digitally recorded and the convolved with an all-pase digital filter which simulates the phase distorting process in tne old recording. The phase of the filter was designed to have a linear ramp phase dispersion with the maximum value of 3 ms at the nyquiet frequency. This distorted speech was used as if it were the only data available with which to etart with the estimation proceee. Proceeeing thie data yields the left hand (first) term of (5).

In order to reveal the second term on the right hand eide of (5), one needs an isolated version of the first term in the right hand eide. To supply this, a second speech passage is recorded by the male eubject number 2 reading from the same text as the previoue recording, and this recording is regarded as the prototype. Processing of this data, than, yields the second term of (5).

The processing yielded the average phase of both recordinge, eo the estimate of $\angle H(f)$ was obtained by subtraction. The aseumption made here ie that the average phase of the prototype recording has the eame mean value as does the original speech to be restored. The estimated and the actual phases (a parabolic curve) of the eyetem are compared in Figure 12.1.

12.5 Discuselon and Future Work

The result is still of little practical applicability for the following reasons: 1) It is not clear what caused the estimated phase of the system to roll off from the actual value in the higher frequency portion as shown in Figure 12.1; 2) Since the length of the data window has to be limited to obtain better unwrapping result, the amount of detectable phase dispersion is also limited to just a few ms. This means the forementioned process will not yield a correct estimate of system phase if this phase dispersion were more than about 5 ms.

The future research will be devoted to finding a way to calculate the reliable unwrapped phase without the restriction of using small windows. If such is shown to be practical, the method will be applied to appropriate test cases.

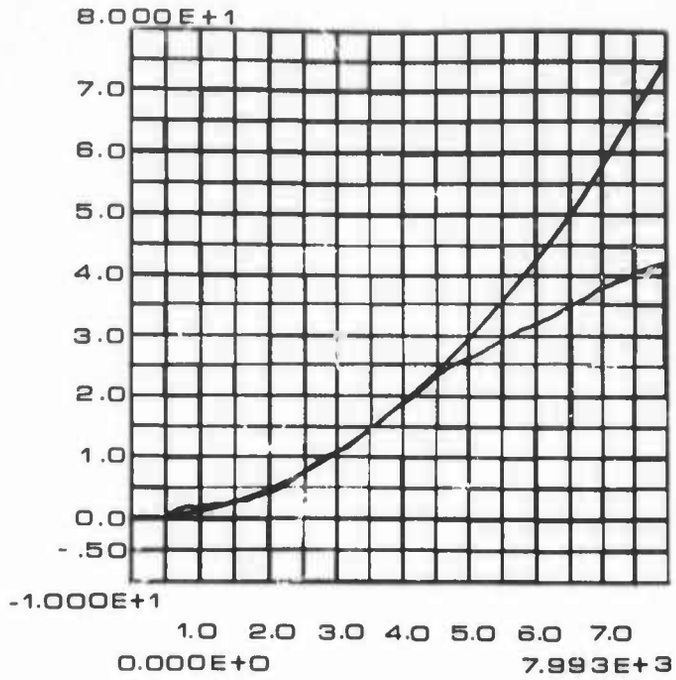


Figure 12.1 -- A comparison of the estimate and the real phase of the linear system. The abcissa is the frequency in Hz, the ordinate is the phase in radians. The continuously increasing parabolic curve is the real phase.

References for Section 12

[1] Goldstein, J.L. "Auditory Spectral Filtering and Monaural Phase Perception." J. Acoust. Soc. Am., Vol. 41, (1967), pp. 458-478.

[2] Mathes, R.C. and R.L. Miller. "Phase Effects in Monaural Perception." J. Acoust. Soc. Am., Vol. 19, p. 780.

[3] Greer, W.H. "Sensitivity of the Ear to Monaural Phase Effects." Ph.D. Thesis, Dept. of Computer Science, University of Utah, Salt Lake City, Utah (1975).

[4] Schafer, R.W. "Echo Removal by Discrete Generalized Linear Filtering." Massachusetts Institute of Technology Technical Report #466, (February 28, 1969), Cambridge, Mass.

[5] Oppenheim, A.v., Schafer, R.W. and T.G. Stockham. "Nonlinear Filtering of Multiplied and Convolved Signals." IEEE Proc., Vol. 56, No. 8, (Aug. 1968), pp. 1264-1291.

[6] Personal Conversation with Dr. R.W. Schafer.

Section 13

Other

Final Computer Science Department technical reports by Rom (UTEC-CSc-75-115) and Ingebretsen (UTEC-CSc-75-118) are in preparation.

A report entitled "Cepstral Prediction Analysis of Digital Waveforms," by Ali Atashroo, was presented at the IEEE 75 Region Six Western Conference. This report describes the continuing attempt to incorporate zeros as well as poles into a spectral model to match a given spectrum. Knowing the impulse train response of a time invariant digital filter with a rational system function, the paper describes a process that identifies the digital filter in terms of the location and order of its poles and zeroes.

PUBLICATIONS AND PRESENTATIONS
Sensory Information Processing

[1] Boll, S.F. "Applications of Linear Predictive Coding to Digital Speech Communication and Recognition." Proc. of the IEEE 1975 Region Six Conference on Communications Technology, (May 1975), Salt Lake City, Utah.

[2] Boll, S.F. "Waveform Comparison Using the Linear Prediction Residual." Computer Science Technical Memorandum #7500, University of Utah, May 1975.

[3] Boll, S.F. Invited tutorial lecture on Linear Predictive Coding given at the National Electronics Conference, Chicago, Ill., October 1975.

[4] Boll, S.F. Invited tutorial lecture on Linear Predictive Coding given at the United States Department of Commerce, Office of Tele- communications, Institute for Telecommunication Sciences, Boulder, Colorado, Feb. 1975.

[5] Boll, S.F. Invited instructor for Short Course given on Digital Speech Analysis and Synthesis presented at the University of Utah, March 1975.

III. RESEARCH ACTIVITIES

SYMBOLIC COMPUTATION

This report summarizes the work of the group during the period of January 1, 1975 through June 30, 1975. The Group's research is directed toward the development of software techniques for the solution of a wide range of symbolic and algebraic problems.

During the last six months, the work of the group has been directed into the following areas:

Section 1

Development of a Mode Analyzing Algebraic Simplification Program

In the last two semi-annual reports we described a new mode analyzing version of REDUCE and our progress towards its implementation. A basic REDUCE system has been completely implemented using these ideas of a complete mode analysis following parsing and only some aspects of the pattern matching, vector and high energy physics packages remain to be checked out before a complete REDUCE system can be released locally for extensive experimentation and testing.

The implementation of a user data definition facility is sufficiently complete to permit all the features of a rather simple type, such as COMPLEX, to be defined quite easily. This includes coercions to and from other types, default values, printing functions, and all the basic operations. A completely satisfactory

definition of a recursive type, such as a polynomial, is almost within our grasp, but still requires some classification of the UNION mode.

We believe that a completely working version of REDUCE, including an integrated data definition facility, will become available during the next six months, and we will then concentrate on rewriting REDUCE using the power of the new system, and extending the data definition mechanism.

Section 2

Research in Independent Compiler and Interpreter Design

The development in this area over the past six months has been particularly exciting. As reported previously, we have extensively rewritten and improved the basic LISP/360 compiler, isolating the code generation into the production of 17 fairly machine independent macros. The latest distributed version of REDUCE for the IBM 360/370 series includes this compiler, written entirely in REDUCE [1]. Recently, a more sophisticated version for the same compiler has been able to produce executable code for the PDP-10, after rewriting the macros. The code produced compares extremely well with the existing LISP 1.6 compiler [2]. In the last few weeks, we have indications that we can produce even better code, and all the optimizations performed should reflect immediately in correspondingly better code for the IBM machines as well. We will in the next few months try this compiler on the 360 to confirm the

effectiveness of the machine independent optimizations, and have also begun an implementation of the macros for the UNIVAC 1108 LISP system.

The work on portable LISP interpreters is progressing satisfactorily, with a complete pseudo-code model now executing on the PDP-10 [3]. This work is being written up as a report, and then an attempt to transfer the program to a different machine will be initiated.

A second approach, using the ideas of abstract machine modeling [4], should ultimately lead to a portable LISP system that can take full advantage of the target machine, and so produce a faster system than the pseudo-code interpreter model. An initial implementation should be completed during the next six month period. The ideal portable LISP system would probably include both pseudo-code (for initial rapid implementation and code packing efficiency), and abstract-machine code (for critical or extremely machine dependent features such as I/O) to produce a system that executes with reasonable speed, and a wide range of capabilities on even quite small machines.

Section 3

Sparse Matrices and Factored Polynomial Algebra

We have continued improving the programs described in the previous reports, with particular emphasis on replacing the expanded polynomial representation by factored polynomials in the different

matrix eolution methods.

The Bareiss (gausaian elimination) method gains significantly in speed, as well as producing a factored result. The Minor Expansion method takes much longer, losing its former advantage. The minor expansion method involves many more algebraic operations constructing many terms that are discarded before the final result. Using the expanded representation, most of these operations involve simpler polynomials than in the Bareiss caee, compensating for the increased number. The factored representation seems to involve a different coet/number tradoff, which we are studying further. One method of improving the minor expansion method is to delay evaluation by constructing a "formal" expreeeion tree. The excess terms will drop out before the final tree is evaluated. Ueing this scheme, we can obtain the result in only slightly more time than the Bareiss method. This "formal" representation has other advantages and we plan to etudy how the advantages of each representation (expanded, factored, formal) can be exploited at the appropriate time in large calculations such as these.

Section 4

Algebraic Applications Packages

A number of extremely useful algebraic packages have been completed and documented within this period. The first is a fairly powerful pattern machine symbolic integrator, capable of handling a very large class of common integrals, with polynomial, rational, trigonomentric and other integrands. This package combinee a number

of interesting techniques, and will provide a "built-in" integration facility previously lacking in REDUCE. The package is easily extensible with the addition of new patterns and will also provide an interface to other more efficient but specialized methods of integration [5].

Another package of some interest is a program for solving a variety of equations in finite terms. This capability is embodied in a general purpose "SOLVE" command, that classifies the equation as whether it has one or more unknowns, whether or not it is non-linear, and whether or not it is simply polynomial, or involves radicals and transcendental functions. By repeated application of algebraic identities, square free factorizations, and reduction of the equation to known linear, quadratic or cubic forms, the program is able to solve a considerable number of quite complex equations. It also provides another convenient interface to the linear equation solving facilities in the system and indicates the need for further extensions, such as the solution of systems of non-linear polynomial equations [6].

The previous two packages were developed as tools to be used in an integral equation package, designed along the line of more common differential equation packages. This package attempts to classify the input equation into one of a number of classes, and then applies a battery of techniques, some exact and some in the form of "analytic" approximations.

While the classification methods may often indicate that an equation will rapidly yield to a particular method (e.g. Laplace Transform, or Separable Kernal), others can only occasionally be solved exactly and some kind of series expansion will at best be able to give some feeling for the behavior of the solution near it s singularities, and dependence on input parameters. Most of the exact methods lead to complicated algebraic equations to solve, while the series methods (Neuman, Taylor, or Fredholm) required repeated analytic integrations or differentiations. In many cases, the package is limited by the current limitations of the SOLVE command or the integration package.

References for Part III

[1] Hearn, A.C. "New Release of REDUCE for IBM System 360/370 Computers." SIGSAM Bulletin, ACM, No. 33 (February 1975) 2.

[2] Quam, L.H. and W. Diffie. "Stanford LISP 1.6 Manual." Stanford Artificial Intelligence Laboratory Operating Note 28.7.

[3] Lugger, J. and H. Melenk. "Darstelling und Bearbeitung ungargreicher LISP--Programme," Angewandte Informatik 6/73, pp. 257-263.

[4] Newey, M.C.; Poole, P.C. and W.M. Waite. "Abstract Machine Modeling to Produce Portable Software--A Review and Evaluation." SOFTWARE-- Practice and Experience, 2 (1972) pp. 107-136.

[5] Risch, R. "The Problem of Integration in Finite Terms." Trans. AMS 139 (1969) pp. 167-189.

[6] Yun, D.Y.Y. "On Algorithms for Solving Systems of Polynomial Equations." SIGSAM Bulletin, ACM, No. 27 (September 1973) pp. 19-25.

PUBLICATIONS AND PRESENTATIONS
Symbolic Computation

[1] Griss, M.L. "REDUCE - A System for Computer Algebra." SIGSAM Fifth Day Session, National Computer Conference, Anaheim, California, May 23, 1975.

[2] Griss, M.L. "REDUCE - A System for Computer Algebra." California Institute of Technology, Los Angeles, May 27, 1975.

[3] Hearn, A.C. "Symbolic Integration Research at the Univereity of Utah." The Los Alamos Workshop on Quadrative, Los Alamos, California, May 16, 1975.

[4] Hearn, A.C. "Scientific Problem Solving by Symbolic Computation." Summer School on Computing Techniques in Physics, Smolenice, Czechoslovakia, June 5-15, 1975.

[5] Stoutemyer, D. "Symbolic Computer Solution of an Equation in Finite Terms." UCP Report No. 33 (1975).

[6] Stoutemyer, D. "Analytical Solution of Integral Equations, Using Computer Algebra." UCP Report No. 34 (1975).

[7] Stoutemyer, D. "A Diminutive REDUCE Program for Symbolic Integration." UCP Report No. 35 (1975).

IV. RESEARCH ACTIVITIES

GRAPHICS

Over the past six months, much of the work of the past several years has been consolidated, focused by the implementation of selected techniques on the new PDP-11/45 based facility. The basis for promising new research efforts involving the results of graphics work has been defined.

Section 1

Complex Scene Image Generation
Martin E. Newell

The properties of procedure models as applied to the representation of three dimensional objects, for the purpose of synthesizing images in the form of shaded pictures, have been investigated. It has been shown that procedure models facilitate the processing of scenes of far greater complexity than has proved practicable using data base modeling techniques alone. The generality and flexibility of procedure models has enabled a system to be implemented which can be, and has been, incrementally expanded to accomodate new model formulations. Examples of output generated by the system are shown in Figures 1.1 and 1.2.

It is believed that the benefits of procedure models are not confined to the field of image synthesis, but have considerable relevance in many areas where modeling of three dimensional objects is of concern, such as computer aided design, computer aided manufacture, stress analysis and dynamics simulation.

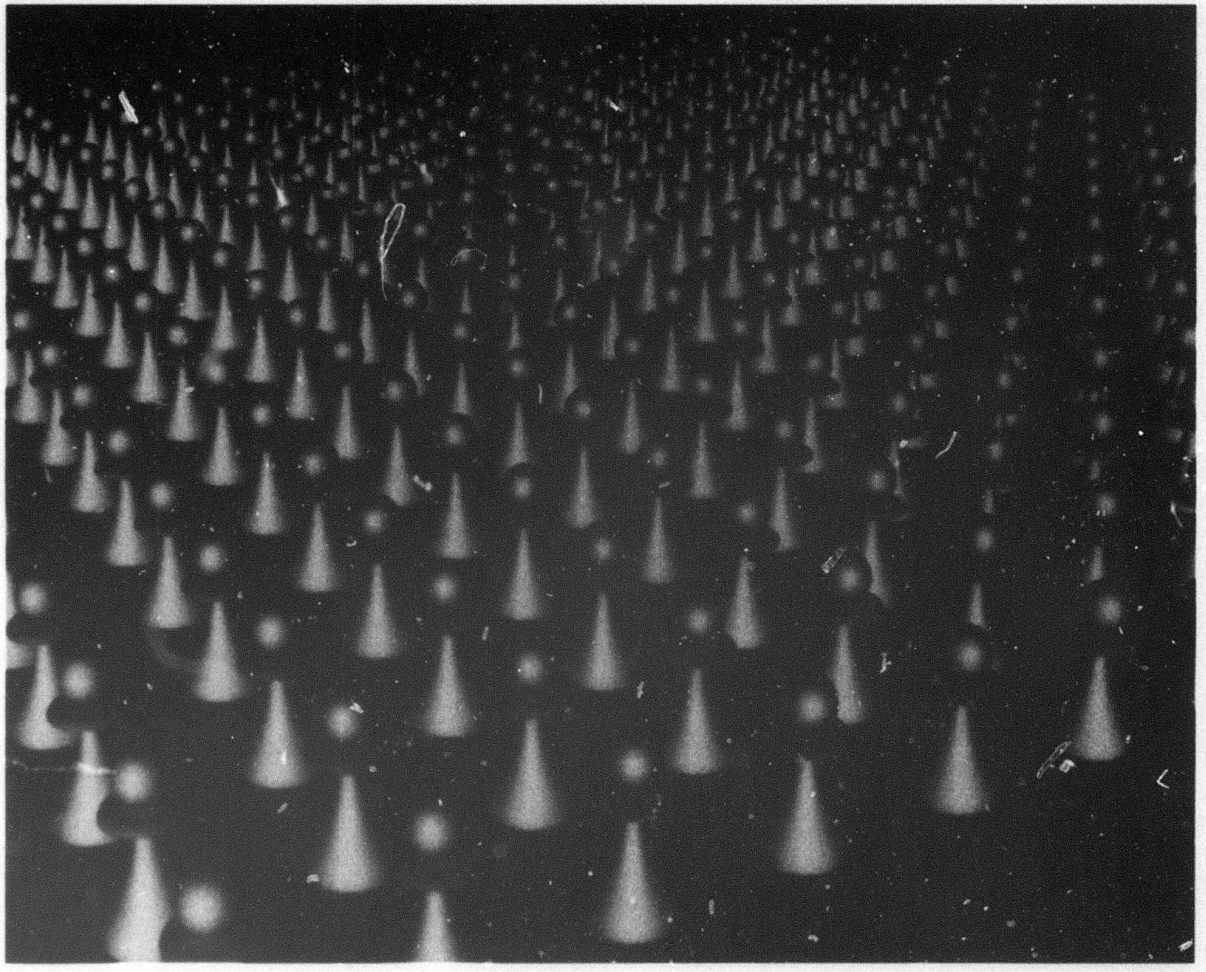


Figure 1.1 -- Image of 361 Pawns generated using Procedural Modelling techniques.

Figure 1.2 -- Scene containing objects using various representations combined using Procedural Modelling Techniques.

Section 2

Measurement and Analysis of 3-D Scenes

Henry Fuchs

Described are the design and implementation of a new range-measuring sensing device and an associated software algorithm for constructing surface descriptions of arbitrary three-dimensional objects from single or multiple views.

The sensing device, which measures surface points from objects in its environment, is a computer-controlled, random-access, triangulating rangefinder with a mirror-deflected laser beam and revolving disc detectors.

The algorithm developed processes these surface points and generates, in a deterministic fashion, complete surface descriptions of all encountered objects. In its processing, the algorithm also detects parts of objects for which there is insufficient data, and can supply the sensing device with the control parameters needed to succesefully measure the uncharted regions.

The resulting object descriptions are suitable for use in a number of areas, such as computer graphics, where the process of constructing object definitions has heretofore been very tedious. Together with the sensing device, this approach to object description can be utilized in a variety of scene analysis and pattern recognition applications which involve interaction with "real world," three-dimensional objects.

Section 3

Real-Time Measurement of 3-D Positions
Larry Evans

Test equipment is currently being acquired and interfaced to the PDP-11/45 computer. It is expected that the summer will see the completion of tests needed to determine the feaelbility and etructure of an advanced position measuring device.

Section 4

Advanced Image Quality
Frank Crow

The problem of aliasing in pictures displayed on a discrete array of picture elements has received continuing attention. A model of the effects of aliasing has been developed. The two main effects, namely loss of small detail and staircase edges, have been almost entirely removed in carefully designed test situations. The application of the developed understanding to exieting visible eurface algorithms is the subject of the efforts during the eummer.

Section 5

PDP-11/45 Facility

Apart from some problems of secondary importance, the frame buffer has been successfully functioning for six months. Only color television output has been used. Problems with the existing precision display have thus far precluded its successful use, but that situation will be resolved in the near future.

PUBLICATIONS AND PRESENTATIONS
Graphics

[1] Catmull, E. "Computer Display of Curved Surfacee." Proc. Conference on Computer Graphics, Pattern Recognition and Data Structure (May 1975).

[2] Clark, J. "Some Properties of B-Splines." Presented at the Second USA-Japan Computer Conference, published in the Proceedings of the Conference, Tokyo, Japan, August 1975.

[3] Crow, F. and B. Tuong-Phong. "Improved Rendition of Polygonal Models of Curved Surfaces." Presented at the Second USA-Japan Computer Conference, published in the Proceedings of the Conference, Tokyo, Japan, August 1975.

[4] Riesenfeld, R.F. "Mathematics of Computer-Aided Geometric Design." Presented at the Dept. of Mathematics, University of Dundee, Scotland.

[5] Riesenfeld, R.F. "Aspects of Modelling in Computer Aided Geometric Design." (A solicited paper) Proc. of NCC, AFIPS Press (1975), pp. 597-602. Presented at the 1975 National Computer Conference, Anaheim, California, May 1975.

[6] Riesenfeld, R.F. "Simulation and Computer Aided Geometric Design." (A solicited paper) Proc. of Society of Photo-Optical Instrumentation Engineers Conference (1975). Presented at the 1975 Conf. of Society of Photo Optical Engineers, Anaheim, California (March 1975).

[7] Riesenfeld, R.F., Dube, R.P. and S.K. Gregory. "Far Out." A 30 second computer generated animation strip with synched sound (1975). To be submitted for exhibition.

[8] Tuong-Phong, B. "Illumination for Computer Generated Pictures." Communications of the ACM, Vol. 18, No. 6 (June 1975).